

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:51:15 ; Search time 222.76 Seconds  
(without alignments)  
1530.574 Million cell updates/sec

Title: US-09-700-770-6  
Perfect score: 543  
Sequence: 1 ccggcgctggggggcgagg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	543	100.0	543	21	AAZ98173 Human signal pepti
2	543	100.0	543	21	AAZ29723 Human lung specifi
3	532.6	98.1	562	19	AAV54620 L0105 specific con
4	529	97.4	570	21	AAZ65103 Membrane-bound pro
5	529	97.4	570	22	AAZ44249 Human PRO1245 (UNQ
6	519	95.6	519	19	AAV54621 L0105 polypeptide
7	241.4	44.5	244	19	AAV54617 L0105 specific pol
8	225	41.4	225	19	AAV54618 L0105 specific pol
9	158	29.1	190	19	AAV54616 L0105 specific pol
10	114	21.0	114	19	AAV54619 L0105 specific pol
11	56.4	10.4	263	19	AAV38067 Human lung tissue

12	56.4	10.4	263	20	AAZ99422 L0103 specific pol
13	56.4	10.4	507	19	AAV38069 Human lung tissue
14	56.4	10.4	507	20	AAZ99424 L0103 specific pol
15	56.4	10.4	507	21	AAZ29721 Human lung specifi
16	56.4	10.4	519	19	AAV38070 Human lung tissue
17	56.4	10.4	519	20	AAZ99425 L0103 specific pol
18	56.4	10.4	531	21	AAZ98174 Human signal pepti
19	50.8	9.4	3957	22	AAA09686 HSV-2 immediate ea
20	50	9.2	58857	21	AAZ58471 Nucleotide sequenc
21	49.6	9.1	357	20	AAV86241 EST clone AA246
22	46.4	8.5	21185	21	AAA63350 Streptomyces globi
23	46.4	8.5	63164	21	AAA63348 Streptomyces globi
24	45.6	8.4	1134	13	AAQ23525 cDNA encoding huma
25	45.6	8.4	6453	19	AAV20436 Human C-H-ras Onco
26	45.6	8.4	6453	21	AAZ88574 Human Ras gene. H
27	45.6	8.4	6453	21	AAZ60602 Nucleotide sequenc
28	45.6	8.4	6453	22	AAZ75128 Human Ras DNA. Ho
29	45.4	8.4	2721	12	AAQ10212 BamHI J-I fragment
30	45.4	8.4	2721	12	AAQ10543 BamHI J-I fragment
31	45.4	8.4	8438	15	AAQ73500 DNA encoding Pseud
32	45.4	8.4	15672	12	AAQ10613 Rianodin receptor
33	44.4	8.2	1209	21	AAZ87293 S. venezuelae deso
34	44.4	8.2	1942	21	AAZ58638 Human PRO1434 prot
35	44.4	8.2	1942	21	AAA51262 Human DNA encoding
36	44.4	8.2	12441	21	AAZ87284 S. venezuelae deso
37	44.4	8.2	13613	21	AAZ87319 S. venezuelae deso
38	44.4	8.2	38506	21	AAZ75633 Nucleotide sequenc
39	44.4	8.2	38506	21	AAZ56001 Recombinant cosmid
40	44	8.1	1133	14	AAZ43306 Human apolipoprote
41	44	8.1	6453	21	AAA48653 Human Ha-ras DNA.
42	44	8.1	6995	16	AAQ85466 Ras oncogene/LacZ
43	44	8.1	7110	22	AAZ81361 Quorum sensing con
44	44	8.1	21034	19	AAV62154 HSV-2 strain SB5 C
45	44	8.1	26338	19	AAV62134 HSV-2 strain SB5 C

ALIGNMENTS

RESULT 1  
AAZ98173  
ID AAZ98173 standard; cDNA; 543 BP.  
XX  
AC AAZ98173;  
XX  
XX  
DT 11-MAY-2000 (first entry)  
XX

DE Human signal peptide containing protein HSP-65 cDNA SEQ ID NO:199.  
KW Human; signal peptide-containing protein; HSP; diagnosis; cancer;  
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
KW antimicrobial; neurotropic; neuroprotective; cardiovascular; hepatotropic;  
KW antischistosomal; gene therapy; cell proliferation; neurological disorder;  
KW reproductive disorder; developmental disorder; arteriosclerosis;  
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
KW Parkinson's disease; Huntington's disease; ovulatory defect;  
KW muscular dystrophy; ss.

OS Homo sapiens.  
XX  
XX WO200000610-A2.  
PN  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-US1484.  
XX  
PR 26-JUN-1998; 98US-0090762.  
PR 31-JUL-1998; 98US-0094983.  
PR 01-OCT-1998; 98US-0102686.  
PR 11-DEC-1998; 98US-0112129.  
XX  
PA (INCY-) INCYTE PHARM INC.

XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;  
 PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;  
 PI Bandman O;  
 XX WPI: 2000-160673/14.  
 DR P-PSDB; AAY87288.  
 XX New human signal peptide-containing proteins useful in treatment,  
 PT prevention and diagnosis of e.g. cancer, inflammation and  
 PT cardiovascular disease  
 XX Claim 9; Page 289; 327pp; English.  
 XX AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the  
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have  
 CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,  
 CC neuroprotective, cardiovascular and antisthmatic activities and can  
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders  
 CC associated with decreased activity or function of HSP. Antagonists of  
 CC HSP are used to treat or prevent disorders associated with increased  
 CC activity or function of HSP. Such diseases include cell proliferation  
 CC (including cancer), inflammation, cardiovascular, neurological,  
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,  
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,  
 CC asthma, Crohn's disease, microbial or other infections, congestive or  
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP  
 CC nucleic acids can be used for the recombinant production of HSP, for  
 CC detecting HSP in standard hybridisation and amplification assays (for  
 CC diagnosis and monitoring), in gene therapy, as antisense,  
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences  
 CC or genetic variations, and for chromosomal mapping. HSP are also used to  
 CC raise specific antibodies (Ab) and to screen for agonists and  
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic  
 CC antagonists, in competitive drug screens, and for purification of HSP  
 CC from natural sources.  
 XX Sequence 543 BP; 89 A; 194 C; 178 G; 82 T; 0 other;

Query Match 100.0%; Score 543; DB 21; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-102;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cggcgctgagggcgagaccgggtataagaagctcgtgcttctagtgagcgcgaagcctgtg 60  
 DB 1 cggcgctgagggcgagaccgggtataagaagctcgtgcttctagtgagcgcgaagcctgtg 60

QY 61 caggttcccgcgcccgagcccgagcccgccatgaagctcgccgctcctctgggctct 120  
 DB 61 caggttcccgcgcccgagcccgagcccgccatgaagctcgccgctcctctgggctct 120

QY 121 gcggtgctctgctcagctcgcgtgcttcttctagtgagcgcgaagcctgtg 180  
 DB 121 gcggtgctctgctcagctcgcgtgcttcttctagtgagcgcgaagcctgtg 180

QY 181 ccagcctgctcgtcgtcgctggagtcggcgagcggagcggcgccgagcctggccaacc 240  
 DB 181 ccagcctgctcgtcgtcgctggagtcggcgagcggagcggcgccgagcctggccaacc 240

QY 241 cctcggaaccctcaaccgctgaagctcgtcgtgagcagcctgagcagcctcggaacc 300  
 DB 241 cctcggaaccctcaaccgctgaagctcgtcgtgagcagcctgagcagcctcggaacc 300

QY 301 acctatagagggctcccaagaagtgctgagctgagctgagctccagcgctggggcg 360  
 DB 301 acctatagagggctcccaagaagtgctgagctgagctgagctccagcgctggggcg 360

QY 361 tgaagccctgaagccctgctggggccctgacagtggttggctgagccgagactggag 420  
 DB 361 tgaagccctgaagccctgctggggccctgacagtggttggctgagccgagactggag 420

QY 421 catctacacctgaggaacagactgcccacccgagggctgaaaccccgcgcggg 480  
 DB 421 catctacacctgaggaacagactgcccacccgagggctgaaaccccgcgcggg 480

QY 481 aggacgctccatcccttccccggccctctcaataaacgtggttaagagcaaaaaa 540  
 DB 481 aggacgctccatcccttccccggccctctcaataaacgtggttaagagcaaaaaa 540

QY 541 aaa 543  
 DB 541 aaa 543

RESULT 2  
 AAZ29723  
 ID AAZ29723 standard; DNA; 543 BP.  
 XX  
 AC AAZ29723;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DE Human lung specific gene Lngl07.  
 XX  
 KW Lung Specific Gene; LSG; Lngl07; human; diagnostic marker;  
 KW prognosticate; lung cancer; diagnosis; ds.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 CDS 93..407  
 FT /\*tag= a  
 FT /product= "LSG Lngl07 protein"  
 FT  
 PN WO960160-A1.  
 XX  
 PD 25-NOV-1999.  
 XX  
 PF 12-MAY-1999; 99WO-US10344.  
 XX  
 PR 21-MAY-1998; 98US-0086212.  
 XX  
 PA (DIAD-) DIADEXUS LLC.  
 XX  
 PI Yang F, Macina RA, Sun Y;  
 XX  
 DR WPI: 2000-116320/10.  
 DR P-PSDB; AAY44458.  
 XX  
 PT A new method for diagnosing, monitoring and staging lung cancer  
 PS Claim 6; Page 36; 40pp; English.  
 XX  
 CC The present sequence is a lung specific gene (LSG) Lngl07 from human  
 CC clone ID 586271. The LSG has high level of tissue specificity for lungs  
 CC and is overexpressed in cancerous tissues. The sequence serves as a  
 CC diagnostic marker for detecting, monitoring, staging and prognosticating  
 CC lung cancer. The diagnosis involves comparing levels of LSG in samples  
 CC obtained from patient and normal control.  
 XX  
 SQ Sequence 543 BP; 89 A; 194 C; 178 G; 82 T; 0 other;

Query Match 100.0%; Score 543; DB 21; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-102;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cggcgctgagggcgagaccgggtataagaagcctcgtgcttggccttggcgggagcg 60  
 DB 1 cggcgctgagggcgagaccgggtataagaagcctcgtgcttggccttggcgggagcg 60

QY 61 caggttcccgcgcccgagcccgagcccgccatgaagctcgccgctcctctgggctct 120  
 DB 61 caggttcccgcgcccgagcccgagcccgccatgaagctcgccgctcctctgggctct 120

QY 121 gcggtgctctgctcagctcgcgtgcttcttctagtgagcgcgaagcctgtg 180  
 DB 121 gcggtgctctgctcagctcgcgtgcttcttctagtgagcgcgaagcctgtg 180

QY 181 ccagcctgctcgtcgtcgctggagtcggcgagcggagcggcgccgagcctggccaacc 240  
 DB 181 ccagcctgctcgtcgtcgctggagtcggcgagcggagcggcgccgagcctggccaacc 240

QY 241 cctcggaaccctcaaccgctgaagctcgtcgtgagcagcctgagcagcctcggaacc 300  
 DB 241 cctcggaaccctcaaccgctgaagctcgtcgtgagcagcctgagcagcctcggaacc 300

QY 301 acctatagagggctcccaagaagtgctgagctgagctgagctccagcgctggggcg 360  
 DB 301 acctatagagggctcccaagaagtgctgagctgagctgagctccagcgctggggcg 360

QY 361 tgaagccctgaagccctgctggggccctgacagtggttggctgagccgagactggag 420  
 DB 361 tgaagccctgaagccctgctggggccctgacagtggttggctgagccgagactggag 420









CC can be used for targeted delivery of bioactive molecules, such as  
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide  
CC sequences, and their fragments, can be used as hybridisation probes, in  
CC chromosomal and gene mapping, and in the generation of anti-sense RNA  
CC and DNA. They may also be used to produce transgenic animals which are  
CC used to develop and screen therapeutically useful reagents. The PRO  
CC nucleotide and protein sequence can be used for tissue typing and in  
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.  
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used  
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and  
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein  
CC sequences given in the exemplification of the present invention.  
XX  
SQ Sequence 570 BP; 129 A; 190 C; 170 G; 81 T; 0 other;

Query Match 97.4%; Score 529; DB 22; Length 570;  
Best Local Similarity 100.0%; Pred. No. 2.4e-99;  
Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 gcagagacgggtataagaagcctgtgacctgcccggcagccgaggttcccccgcgc 74  
DB 1 gcgagacgggtataagaagcctgtgacctgcccggcagccgaggttcccccgcgc 60  
QY 75 gccccgagccccgcgcacatgaagctgcgcgcctcctcctgggctgtgctgcccctgtcc 134  
DB 61 gccccgagccccgcgcacatgaagctgcgcgcctcctcctgggctgtgctgcccctgtcc 120  
QY 135 tgcagctccgctgtcttcttagtggtctgcccagccttgcccagcctgtcgtcgt 194  
DB 121 tgcagctccgctgtcttcttagtggtctgcccagccttgcccagcctgtcgtcgtcgt 180  
QY 195 gccttgagctgc 254  
DB 181 gccttgagctgc 240  
QY 255 aaccgcgtgaagctcctgtgagcagcctgggctatccctggcagcctcatagagggc 314  
DB 241 aaccgcgtgaagctcctgtgagcagcctgggctatccctggcagcctcatagagggc 300  
QY 315 tccagaagctgtggtgagctgggtcccccagcgcgtggggccctgaagccctgaag 374  
DB 301 tccagaagctgtggtgagctgggtcccccagcgcgtggggccctgaagccctgaag 360  
QY 375 gcctctggtgggccccctgcagctgttggctgagccgagactggagcattacacctgag 434  
DB 361 gcctctggtgggccccctgcagctgttggctgagccgagactggagcattacacctgag 420  
QY 435 gacaaagcgtgcccacccgcgagggctgaaacccccgcgcgcgcgcgcgcgcgcgcgcgc 494  
DB 421 gacaaagcgtgcccacccgcgagggctgaaacccccgcgcgcgcgcgcgcgcgcgcgcgc 480  
QY 495 ccttccccgcgcctcctaataacgtgttaagacaaaataaaaaa 543  
DB 481 ccttccccgcgcctcctaataacgtgttaagacaaaataaaaaa 529

RESULT 6

AAV54621  
ID AAV54621 standard; CDNA; 519 BP.  
XX  
AC AAV54621;  
XX  
DT 30-OCT-1998 (first entry)  
XX  
DE LU105 polypeptide encoding CDNA clone 13278361H.  
XX  
KW LU105; lung disease marker; immunoassay; lung disease; cancer;  
KW blood; plasma; serum; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers

FT CDS 79..393  
FT /\*tag= a  
FT /transl\_except= (pos:136..138, aa:Val)  
FT /product= "LU105 polypeptide"  
XX  
PN WO9833926-A1.  
XX  
PD 06-AUG-1998.  
XX  
PF 30-JAN-1998; 98WO-US01766.  
XX  
PR 31-JAN-1997; 97US-0791710.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;  
PI Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
DR WPI; 1998-437479/37.  
DR P-PSDB; AAW75868.  
XX  
PT New nucleic acid for the lung disease marker LU105 - polypeptides,  
PT antibodies and genes, used for diagnosis, prevention, treatment of  
PT lung disease, specifically cancer  
XX  
PS Claim 11; Fig 1; 123pp; English.  
XX  
CC Sequences shown in AAV54616 to AAV54621 represent LU105 specific  
CC polynucleotide sequences. These are used in the method of the invention  
CC for detecting target LU105 nucleic acid. The method comprises treating a  
CC sample with at least one LU105 specific nucleic acid, or its complement  
CC which is at least 50 percent identical with the LU105 specific nucleic  
CC acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.  
CC Cells transformed with a recombinant expression system that contains  
CC LU105 specific nucleic acid fragments, are used to express recombinant  
CC LU105 polypeptides which are used to raise antibodies. The antibodies are  
CC used to detect the LU105 antigen, and correspondingly this antigen is  
CC used to detect specific antibodies, in usual immunoassays. The LU105  
CC polypeptides and nucleic acid sequences are used for diagnosis, staging,  
CC monitoring, prognosis, prevention, treatment and determination of  
CC susceptibility to, lung disease, specifically cancer. The LU105  
CC polypeptides are also used to screen for specific binding agents, useful  
CC therapeutically. LU105 is a marker for lung disease (present at high  
CC concentration, in altered form or in an unusual body compartment). LU105  
CC can be detected in blood, plasma or serum in an inexpensive, non-invasive  
CC test.  
XX  
SQ Sequence 519 BP; 78 A; 190 C; 170 G; 81 T; 0 other;

Query Match 95.6%; Score 519; DB 19; Length 519;  
Best Local Similarity 100.0%; Pred. No. 2.6e-97;  
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 gcagagacgggtataagaagcctgtgacctgcccggcagccgaggttcccccgcgc 74  
DB 1 gcgagacgggtataagaagcctgtgacctgcccggcagccgaggttcccccgcgc 60  
QY 75 gccccgagccccgcgcacatgaagctgcgcgcctcctcctgggctgtgctgcccctgtcc 134  
DB 61 gccccgagccccgcgcacatgaagctgcgcgcctcctcctgggctgtgctgcccctgtcc 120  
QY 135 tgcagctccgctgtcttcttagtggtctgcccagccttgcccagcctgtcgtcgt 194  
DB 121 tgcagctccgctgtcttcttagtggtctgcccagccttgcccagcctgtcgtcgtcgt 180  
QY 195 gccttgagctgc 254  
DB 181 gccttgagctgc 240  
QY 255 aaccgcgtgaagctcctgtgagcagcctgggctatccctggcagcctcatagagggc 314  
DB 241 aaccgcgtgaagctcctgtgagcagcctgggctatccctggcagcctcatagagggc 300

Db 241 aaccgctgaagctcgtgagcagcctgggcatcccgctgaaccacctcatagagggc 300  
Qy 315 tcccaagaagtgtgctgagctgggtcccccagccgctggggccgtgaagccctgaag 374  
Db 301 tcccaagaagtgtgctgagcctgggtcccccagccgctggggccgtgaagccctgaag 360  
Qy 375 gccctgctggggccctgacagctgttggctgagccgagactggagctatcacactgag 434  
Db 361 gccctgctggggccctgacagctgttggctgagccgagactggagctatcacactgag 420  
Qy 435 gacaaagacgtgccccacccgagggctgaaacccccccgagggagacgctccatcc 494  
Db 421 gacaaagacgtgccccacccgagggctgaaacccccccgagggagacgctccatcc 480  
Qy 495 ccttccccggccctctcaataaacgtggttaagagca 533  
Db 481 ccttccccggccctctcaataaacgtggttaagagca 519  
  
RESULT 7  
AAV54617  
ID AAV54617 standard; cDNA; 244 BP.  
XX  
AC AAV54617;  
XX  
DT 30-OCT-1998 (first entry)  
XX  
DE LU105 specific polynucleotide sequence from clone 1327836.  
XX  
KW LU105; lung disease marker; immunoassay; lung disease; cancer;  
KW blood; plasma; serum; ss.  
OS Homo sapiens.  
XX  
PN W09833926-A1.  
XX  
PD 06-AUG-1998.  
XX  
PF 30-JAN-1998; 98WO-US01766.  
XX  
PR 31-JAN-1997; 97US-0791710.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;  
PI Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
DR WPI; 1998-437479/37.  
XX  
XX New nucleic acid for the lung disease marker LU105 - polypeptides,  
XX antibodies and genes, used for diagnosis, prevention, treatment of  
XX lung disease, specifically cancer  
PS Claim 11; Fig 1; 123pp; English.  
XX  
XX Sequences shown in AAV54616 to AAV54621 represent LU105 specific  
XX polynucleotide sequences. These are used in the method of the invention  
XX for detecting target LU105 nucleic acid. The method comprises treating a  
XX sample with at least one LU105 specific nucleic acid, or its complement  
XX which is at least 50 percent identical with the LU105 specific nucleic  
XX acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.  
XX Cells transformed with a recombinant expression system that contains  
XX LU105 specific nucleic acid fragments, are used to express recombinant  
XX LU105 polypeptides which are used to raise antibodies. The antibodies are  
XX used to detect the LU105 antigen, and correspondingly this antigen is  
XX used to detect specific antibodies, in usual immunoassays. The LU105  
XX polypeptides and nucleic acid sequences are used for diagnosis, staging,  
XX monitoring, prognosis, prevention, treatment and determination of  
XX susceptibility to, lung disease, specifically cancer. The LU105  
XX polypeptides are also used to screen for specific binding agents, useful  
XX therapeutically. LU105 is a marker for lung disease (present at high  
XX concentration, in altered form or in an unusual body compartment). LU105

CC can be detected in blood, plasma or serum in an inexpensive, non-invasive  
CC test.  
XX  
SQ Sequence 244 BP; 25 A; 97 C; 82 G; 39 T; 1 other;  
  
Query Match 44.5%; Score 241.4; DB 19; Length 244;  
Best Local Similarity 99.2%; Pred. No. 6.9e-41;  
Matches 242; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 15 gcgagagccgggtataaagaacctgctgctgcccgggagccgaggttccccgcgc 74  
Db 1 gcgagagccgggtataaagaacctgctgctgcccgggagccgaggttccccgcgc 60  
Qy 75 gccccgagccccgcgccataaagctccgccctccctgggctctgctgagccctgtcc 134  
Db 61 gccccgagccccgcgccatgaagctccgccctccctgggctgctgctgctgctcc 120  
Qy 135 tgcagctccgctgctgcttctttagtggtgctggcgccaaagcctgtgcccagcctgtcgt 194  
Db 121 tgcagctccgctgctgcttctttagtggtgctggcgccaaagcctgtgcccagcctgtcgt 180  
Qy 195 gcgctggagtcgagcgccgagggccgagggccgagccctggccaaacccctcggaaccctc 254  
Db 181 gcgctggagtcgagcgccgagggccgagggccgagccctggccaaacccctcggaaccctc 240  
Qy 255 aacc 258  
Db 241 aacc 244  
  
RESULT 8  
AAV54618  
ID AAV54618 standard; cDNA; 225 BP.  
XX  
AC AAV54618;  
XX  
DT 30-OCT-1998 (first entry)  
XX  
DE LU105 specific polynucleotide sequence from clone 1605935.  
XX  
KW LU105; lung disease marker; immunoassay; lung disease; cancer;  
KW blood; plasma; serum; ss.  
OS Homo sapiens.  
XX  
PN W09833926-A1.  
XX  
PD 06-AUG-1998.  
XX  
PF 30-JAN-1998; 98WO-US01766.  
XX  
PR 31-JAN-1997; 97US-0791710.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;  
PI Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
DR WPI; 1998-437479/37.  
XX  
XX New nucleic acid for the lung disease marker LU105 - polypeptides,  
XX antibodies and genes, used for diagnosis, prevention, treatment of  
XX lung disease, specifically cancer  
PS Claim 11; Fig 1; 123pp; English.  
XX  
XX Sequences shown in AAV54616 to AAV54621 represent LU105 specific  
XX polynucleotide sequences. These are used in the method of the invention  
XX for detecting target LU105 nucleic acid. The method comprises treating a  
XX sample with at least one LU105 specific nucleic acid, or its complement  
XX which is at least 50 percent identical with the LU105 specific nucleic  
XX acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.  
XX Cells transformed with a recombinant expression system that contains  
XX LU105 specific nucleic acid fragments, are used to express recombinant  
XX LU105 polypeptides which are used to raise antibodies. The antibodies are  
XX used to detect the LU105 antigen, and correspondingly this antigen is  
XX used to detect specific antibodies, in usual immunoassays. The LU105  
XX polypeptides and nucleic acid sequences are used for diagnosis, staging,  
XX monitoring, prognosis, prevention, treatment and determination of  
XX susceptibility to, lung disease, specifically cancer. The LU105  
XX polypeptides are also used to screen for specific binding agents, useful  
XX therapeutically. LU105 is a marker for lung disease (present at high  
XX concentration, in altered form or in an unusual body compartment). LU105

RESULT	9	
AAV54616		
ID	AAV54616	standard; cDNA; 190 BP.
XX		
AC	AAV54616;	
XX		
DT	30-OCT-1998	(first entry)
XX		
DE	LUI05	specific polynucleotide sequence from clone 3353867.
XX		
XX	LUI05;	lung disease marker; immunoassay; lung disease; cancer;
KW	blood; plasma; serum; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	W09833926-A1.	
XX		
PD	06-AUG-1998.	
XX		
PF	30-JAN-1998;	98WO-US01766.
XX		
PR	31-JAN-1997;	97US-0791710.
XX		
PA	(ABBO )	ABBOTT LAB.
XX		
PI	Billington-Medel PA,	Cohen M, Colpitts TL, Friedman PN;
PI	Gordon J, Granadosen,	Hodges SC, Klass MR, Kratochvil JD;
PI	Roberts-Rapp L, Russell JC,	Stroupe SD;
XX		
DR	WPI;	1998-437479/37.
XX		
PT	New nucleic acid for the lung disease marker LUI05 - polypeptides,	
PT	antibodies and genes, used for diagnosis, prevention, treatment of	
PT	lung disease, specifically cancer	

RESULT	10	
AAV54619		
ID	AAV54619	standard; cDNA; 114 BP.
XX		
AC	AAV54619;	
XX		
DT	30-OCT-1998	(first entry)
XX		
DE	LU105	specific polynucleotide sequence from clone 811640.
XX		
KW	LU105;	lung disease marker; immunoassay; lung disease; cancer;
KW	blood; plasma;	serum; ss.
XX		
OS	Homo sapiens.	
XX		
PN	WO9833926-A1.	
XX		
PD	06-AUG-1998.	
XX		
PF	30-JAN-1998;	98WO-US01766.
XX		
PR	31-JAN-1997;	97US-0791710.
XX		
PA	(ABBO )	ABBOTT LAB.
XX		
PI	Billing-Medel PA,	Cohen M, Colpitts TL, Friedman PN;
PI	Gordon J, Granadosen,	Hodges SC, Klass MR, Kratochvil JD;
PI	Roberts-Rapp L,	Russell JC, Stroupe SD;
XX		
DR	WPI;	1998-437479/37.

XX New nucleic acid for the lung disease marker LU105 - polypeptides,  
 PT antibodies and genes, used for diagnosis, prevention, treatment of  
 PT lung disease, specifically cancer  
 XX  
 PS Claim 11; Fig 1; 123pp; English.  
 XX  
 CC Sequences shown in AAV54616 to AAV54621 represent LU105 specific  
 CC polynucleotide sequences. These are used in the method of the invention  
 CC for detecting target LU105 nucleic acid. The method comprises treating a  
 CC sample with at least one LU105 specific nucleic acid, or its complement  
 CC which is at least 50 percent identical with the LU105 specific nucleic  
 CC acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.  
 CC Cells transformed with a recombinant expression system that contains  
 CC LU105 specific nucleic acid fragments, are used to express recombinant  
 CC LU105 polypeptides which are used to raise antibodies. The antibodies are  
 CC used to detect the LU105 antigen, and correspondingly this antigen is  
 CC used to detect specific antibodies, in usual immunoassays. The LU105  
 CC polypeptides and nucleic acid sequences are used for diagnosis, staging,  
 CC monitoring, prognosis, prevention, treatment and determination of  
 CC susceptibility to, lung disease, specifically cancer. The LU105  
 CC polypeptides are also used to screen for specific binding agents, useful  
 CC therapeutically. LU105 is a marker for lung disease (present at high  
 CC concentration, in altered form or in an unusual body compartment). LU105  
 CC can be detected in blood, plasma or serum in an inexpensive, non-invasive  
 CC test.  
 XX  
 SQ Sequence 114 BP; 26 A; 44 C; 29 G; 15 T; 0 other;

Query Match 21.0%; Score 114; DB 19; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-15;  
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy - 419 agcatctacactgagacagacgctgccaccgcgagggctgaaaccccgccg 478  
 Db 1 agcatctacactgagacagacgctgccaccgcgagggctgaaaccccgccg 60  
 Qy - 479 ggagaccgtccatccctcccgccgcccctctcaataaacggtggttaagac 532  
 Db 61 ggagaccgtccatccctcccgccgcccctctcaataaacggtggttaagac 114

RESULT 11  
 AAV38067  
 ID AAV38067 standard; cDNA; 263 BP.  
 AC AAV38067;  
 XX  
 DT 14-SEP-1998 (first entry)  
 XX  
 DE Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:2.  
 XX  
 KW Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9820143-A1.  
 XX  
 PD 14-MAY-1998.  
 XX  
 PF 05-NOV-1997; 97WO-US20680.  
 XX  
 PR 05-NOV-1996; 96US-0744211.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
 XX  
 DR WPI; 1998-286957/25.  
 XX  
 PT Lung tissue derived polynucleotide LU103 - useful to detect,  
 XX

PT diagnose, stage, monitor, prognosis, prevent, treat or determine  
 PT pre-disposition to lung disease, e.g. lung cancer  
 XX  
 PS Claim 1; Page 67; 86pp; English.  
 XX  
 CC The present sequence represents a polynucleotide specific for lung  
 CC tissue gene LU103. A method has been developed for detecting the  
 CC presence of a target LU103 polynucleotide in a test sample, comprising:  
 CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,  
 CC and (b) detecting the target LU103 polynucleotide in the test sample,  
 CC where the LU103 polynucleotide has at least 50% identity to the 263,  
 CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to  
 CC AAV38070. The methods and products of the present invention may be used  
 CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or  
 CC determine the predisposition diseases and conditions of the lung, e.g.  
 CC lung cancer.  
 XX  
 SQ Sequence 263 BP; 64 A; 57 C; 71 G; 71 T; 0 other;  
 Query Match 10.4%; Score 56.4; DB 19; Length 263;  
 Best Local Similarity 60.4%; Pred. No. 0.0028;  
 Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
 Qy 229 cccctggccaccccccctggccaccctcaaccgcgctgaagctcgtgagcagcctgggca 288  
 Db 60 ctctggacaacattctctccctttatggtatccattaaagcttcttctgaaactctgggca 119  
 Qy 289 tccccgtgaaccacctcatagaggctcccgagagctgtgtgctgagctgggtcccccagg 348  
 Db 120 ttctgttgagcaccctgtggagggttaaggaagtgttaaatgagctgggaccagagg 179  
 Qy 349 ccgtggggccgctgaaggccctgaagccctgct 382  
 Db 180 ctctgaagctgtgaagaaactgctggaggcgct 213  
 RESULT 12  
 AAX99422  
 ID AAX99422 standard; DNA; 263 BP.  
 XX  
 AC AAX99422;  
 XX  
 DT 19-OCT-1999 (first entry)  
 XX  
 DE LU103 specific polynucleotide isolated from clone 1235531.  
 XX  
 KW LU103; tumour; lung cancer; detection; FISH;  
 KW Fluorescent in situ hybridisation; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN US593265-A.  
 XX  
 PD 17-AUG-1999.  
 XX  
 PF 05-NOV-1997; 97US-0964725.  
 XX  
 PR 05-NOV-1997; 97US-0964725.  
 XX  
 PR 05-NOV-1996; 96US-0744211.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
 XX  
 DR WPI; 1999-468402/39.  
 XX  
 PT Polynucleotides useful for detecting, diagnosing and monitoring  
 PT diseases of the lung such as lung cancer  
 XX  
 PS Claim 1; Column 47; 36pp; English.

XX The 263 base pair sequence of a LU103 specific polynucleotide was  
CC derived from clone 1235531.  
CC The polynucleotides and methods disclosed in the invention can be  
CC useful for detecting, diagnosing, staging, monitoring or predicting  
CC diseases and conditions of the lung, such as lung cancer.  
CC The polynucleotides may be used to produce probes for use in  
CC fluorescent in situ hybridization (FISH) technology to perform  
CC chromosomal analysis and identify cancer specific alterations  
CC such as deletions.  
XX  
SQ Sequence 263 BP; 64 A; 57 C; 71 G; 71 T; 0 other;  
  
Query Match 10.4%; Score 56.4; DB 20; Length 263;  
Best Local Similarity 60.4%; Pred. No. 0.0028;  
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
  
QY 229 cctgtgcaacccctcgacccctcaaccgctgaagctcctgctgagcagctgggca 288  
DB 60 ctctggacaacattctccctttatggtatccattaaagcttcttgaacctctgggca 119  
  
QY 289 tcccggtgaacacctcatagagggtcccgaaagtgtgtggtgagctgggtcccccagg 348  
DB 120 ttctgttgagcaccctgtggagggtctaaagaaagtgtgtaaatgagctgggaccagagg 179  
  
QY 349 ccgtggggccgctgaagccctgaagccctgct 382  
DB 180 ctctgaagctgtgaagaacctgctggaggcgt 213  
  
RESULT 13  
AAV38069  
ID AAV38069 standard; cDNA; 507 BP.  
XX  
AC AAV38069;  
XX  
DT 14-SEP-1998 (first entry)  
XX  
DE Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:4.  
XX  
KW Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 83..364  
FT /\*tag= a  
FT /product= "LU103"  
XX  
PN W09820143-A1.  
XX  
PD 14-MAY-1998.  
XX  
PF 05-NOV-1997; 97WO-US20680.  
XX  
PR 05-NOV-1996; 96US-0744211.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
PI Kratochvill JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
DR WPI; 1998-286957/25.  
DR P-PSDB; AAW62068.  
XX  
PT Lung tissue derived polynucleotide LU103 - useful to detect,  
PT diagnose, stage, monitor, prognosis, prevent, treat or determine  
PT pre-disposition to lung disease, e.g. lung cancer  
XX  
XX Claim 1; Page 67; 86pp; English.  
XX  
XX The present sequence represents a polynucleotide specific for lung

CC tissue gene LU103. A method has been developed for detecting the  
CC presence of a target LU103 polynucleotide in a test sample, comprising:  
CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,  
CC and (b) detecting the target LU103 polynucleotide in the test sample,  
CC where the LU103 polynucleotide has at least 50% identity to the 269,  
CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to  
CC AAV38070. The methods and products of the present invention may be used  
CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or  
CC determine the predisposition diseases and conditions of the lung, e.g.  
CC lung cancer.  
XX  
SQ Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;  
  
Query Match 10.4%; Score 56.4; DB 19; Length 507;  
Best Local Similarity 60.4%; Pred. No. 0.0029;  
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
  
QY 229 cctgtgcaacccctcgacccctcaaccgctgaagctcctgctgagcagctgggca 288  
DB 195 ctctggacaacattctccctttatggtatccattaaagcttcttgaacctctgggca 254  
  
QY 289 tcccggtgaacacctcatagagggtcccgaaagtgtgtggtgagctgggtcccccagg 348  
DB 255 ttctgttgagcaccctgtggagggtctaaagaaagtgtgtaaatgagctgggaccagagg 314  
  
QY 349 ccgtggggccgctgaagccctgaagccctgct 382  
DB 315 ctctgaagctgtgaagaacctgctggaggcgt 348  
  
RESULT 14  
AAV99424  
ID AAV99424 standard; DNA; 507 BP.  
XX  
AC AAV99424;  
XX  
DT 19-OCT-1999 (first entry)  
XX  
DE LU103 specific polynucleotide consensus sequence.  
XX  
KW LU103; tumour; lung cancer; detection;  
KW Fluorescent in situ hybridisation; ss.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 85..364  
FT /\*tag= a  
FT /product= "Sequence AAV28334 amino acid residue"  
XX  
PN US5939265-A.  
XX  
PD 17-AUG-1999.  
XX  
PF 05-NOV-1997; 97US-0964725.  
XX  
PR 05-NOV-1997; 97US-0964725.  
PR 05-NOV-1996; 96US-0744211.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
PI Kratochvill JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
DR WPI; 1999-468402/39.  
DR P-PSDB; AAV28334.  
XX  
PT Polynucleotides useful for detecting, diagnosing and monitoring  
PT diseases of the lung such as lung cancer  
XX  
XX Claim 1; Column 47-49; 36pp; English.

XX The 507 base pair sequence of a Lng103 specific polynucleotide was  
 CC derived from a consensus of the isolated sequences from clones  
 CC 1235095, 1235531 and 1379417.  
 CC The polynucleotides and methods disclosed in the invention can be  
 CC useful for detecting, diagnosing, staging, monitoring or predicting  
 CC diseases and conditions of the lung, such as lung cancer.  
 CC The polynucleotides may be used to produce probes for use in  
 CC fluorescent in situ hybridization (FISH) technology to perform  
 CC chromosomal analysis and identify cancer specific alterations  
 CC such as deletions.  
 XX  
 SQ Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;  
  
 Query Match 10.4%; Score 56.4; DB 20; Length 507;  
 Best Local Similarity 60.4%; Pred. No. 0.0029;  
 Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
  
 Qy 229 ccctggccacccctcggcacccctcaaccgctgaagctcctgctgagcagcctgggca 288  
 Db 195 ctctggacaacattctccctttatgatccattaaagcttctctgaaaaactctgggca 254  
 Qy 289 tcccgtgaaccacctcatagagggctcccagaagtgctggtgagctgggtcccccag 348  
 Db 255 ttctgttgagcaccctgtggaggggctaaagagtggtataatgagctgggaccagagg 314  
 Qy 349 ccgtgggggcccgtgaagccctgaagccctgct 382  
 Db 315 ctctgaagctgtgaagaaactgctggaggcgt 348  
  
 RESULT 15  
 AAZ29721  
 ID AAZ29721 standard; DNA; 507 BP.  
 XX  
 AC AAZ29721;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DE Human lung specific gene Lng101.  
 XX  
 KW Lung Specific Gene; LSG; Lng101; human; diagnostic marker;  
 KW prognosticate; lung cancer; diagnosis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 83..364  
 FT /\*tag= a  
 FT /product= "LSG Lng101 protein"  
 XX  
 PN WO9960160-A1.  
 XX  
 XX 25-NOV-1999.  
 XX  
 XX 12-MAY-1999; 99WO-US10344.  
 XX  
 XX 21-MAY-1998; 98US-0086212.  
 XX  
 XX (DIAD-) DIADEXUS LLC.  
 XX  
 XX Yang F, Macina RA, Sun Y;  
 XX  
 XX WPI; 2000-116320/10.  
 DR P-PSDB; AAY4456.  
 XX  
 XX A new method for diagnosing, monitoring and staging lung cancer  
 XX  
 XX Claim 6; Page 33; 40pp; English.  
 XX  
 CC The present sequence is a lung specific gene (LSG) Lng101 from human  
 CC clone ID 126758. The LSG has high level of tissue specificity for lungs

CC and is overexpressed in cancerous tissues. The sequence serves as a  
 CC diagnostic marker for detecting, monitoring, staging and prognosticating  
 CC lung cancer. The diagnosis involves comparing levels of LSG in samples  
 CC obtained from patient and normal control.  
 XX  
 SQ Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;  
  
 Query Match 10.4%; Score 56.4; DB 21; Length 507;  
 Best Local Similarity 60.4%; Pred. No. 0.0029;  
 Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
  
 Qy 229 ccctggccacccctcggcacccctcaaccgctgaagctcctgctgagcagcctgggca 288  
 Db 195 ctctggacaacattctccctttatgatccattaaagcttctctgaaaaactctgggca 254  
 Qy 289 tcccgtgaaccacctcatagagggctcccagaagtgctggtgagctgggtcccccag 348  
 Db 255 ttctgttgagcaccctgtggaggggctaaagagtggtataatgagctgggaccagagg 314  
 Qy 349 ccgtgggggcccgtgaagccctgaagccctgct 382  
 Db 315 ctctgaagctgtgaagaaactgctggaggcgt 348  
  
 Search completed: November 17, 2001, 14:51:18  
 Job time: 7206 sec

THIS PAGE BLANK (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:47:27 ; Search time 115.94 Seconds  
(without alignments)  
4024.018 Million cell updates/sec

Title: US-09-700-770-3  
Perfect score: 2060  
Sequence: 1 cttgagactctcaataact.....ttccattgagaaaaa 2060

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued\_Patents\_NA.\*
- 1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*
  - 2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*
  - 3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*
  - 4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*
  - 5: /cgn2.6/ptodata/1/ina/PTUTS\_COMB.seq.\*
  - 6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2060	100.0	2061	4	US-09-008-271A-16
2	1356.8	65.9	1910	4	US-08-974-691-7
3	1199.2	58.2	1299	1	US-08-723-938-4
4	1199.2	58.2	1299	2	US-09-080-538-4
5	1142	55.4	1353	4	US-08-974-691-1
6	747	36.3	1651	4	US-08-974-691-5
7	316.6	15.4	2038	2	US-08-631-097-7
8	316.6	15.4	2038	4	US-08-810-712-11
9	316.6	15.4	2465	5	PCT-US92-08090-1
10	168.2	8.2	1240	1	US-08-240-372-2
11	168.2	8.2	2073	4	US-09-032-523-6
12	168.2	8.2	2733	2	US-08-846-021A-6
13	167.8	8.1	2732	6	5217891-14
14	151.8	7.4	2875	1	US-08-328-314-1
15	151.8	7.4	2875	1	US-08-731-045-1
16	139.2	6.8	2454	3	US-09-079-415-1
17	133.6	6.5	1615	1	US-08-360-673-5
18	123.6	6.0	2032	1	US-08-088-633-1
19	123.6	6.0	2032	1	US-08-245-756-1
20	123.6	6.0	2032	1	US-08-441-750-1
21	123.6	6.0	2032	1	US-08-441-751-1
22	123.6	6.0	2032	5	PCT-US92-02521-1
23	88.4	4.3	1800	3	US-09-039-773A-1
24	73	3.5	1245	4	US-09-318-443-1
25	69.4	3.4	1682	4	US-09-318-443-7
26	68.2	3.3	43676	3	US-09-356-952-12
27	67.8	3.3	1536	4	US-09-318-443-5

28	62.8	3.0	1867	2	US-08-607-509-3	Sequence 3, Appli
29	62.8	3.0	1867	2	US-08-634-642-3	Sequence 3, Appli
30	62.8	3.0	1867	3	US-08-989-370-3	Sequence 3, Appli
31	61.8	3.0	2934	3	US-09-149-934-2	Sequence 2, Appli
32	59	2.9	1254	5	PCT-US96-05320A-894	Sequence 894, App
33	58.2	2.8	2365	4	US-09-183-706-42	Sequence 42, Appl
34	58.2	2.8	2365	4	US-09-567-995-42	Sequence 42, Appl
35	58	2.8	1618	2	US-08-533-669A-9	Sequence 9, Appli
36	58	2.8	1618	2	US-08-607-509-1	Sequence 1, Appli
37	58	2.8	1618	2	US-08-454-036-1	Sequence 1, Appli
38	58	2.8	1618	2	US-08-634-642-1	Sequence 1, Appli
39	58	2.8	1618	3	US-08-989-370-1	Sequence 1, Appli
40	58	2.8	1618	5	PCT-US95-05064-1	Sequence 1, Appli
41	57.2	2.8	1374	3	US-08-929-738-1	Sequence 1, Appli
42	57.2	2.8	1614	3	US-08-929-738-2	Sequence 2, Appli
43	56.8	2.8	3825	4	US-09-208-742-3	Sequence 3, Appli
44	55.6	2.7	3408	3	US-09-058-489-14	Sequence 14, Appli
45	55.6	2.7	5322	3	US-09-058-489-13	Sequence 13, Appli

ALIGNMENTS

RESULT 1  
US-09-008-271A-16  
Sequence 16, Application US/09008271A  
Patent No. 6203979  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
Hillman, Jennifer L.  
Yue, Henry  
Guegler, Karl J.  
Corley, Neil C.  
Tang, Tom Y.  
Shah, Purvi  
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/008,271A  
FILING DATE: 16-Jan-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mohan-Peterson, Sheela  
REGISTRATION NUMBER: 41,201  
REFERENCE/DOCKET NUMBER: PF-0458 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2061 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LUNGAST01  
CLONE: 877617  
SEQUENCE DESCRIPTION: SEQ ID NO: 16 :  
US-09-008-271A-16

Query Match. 100.0%; Score 2060; DB 4; Length 2061;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2060; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	cttgaagactcctaaactgtgtcatggtgaagccgaccgaaataactgaatgattt	60
Db	1	CTTGAGAGCTCTCAAAATCTTGTCATGGATGAAGCGACCGAATCTGAATATGGATTT	60
Qy	61	tgagacagaggttgacaagcctcgagatcggaataacttctctctgtccaccatgac	120
Db	61	TGAGACAGAGTTTGACAAGCTTCGAGATCGAATACTCGAAATCTCTCTGCCACCATGAC	120
Qy	121	caagaaggttcaaaaactcagcgagcagctctgaagaatcctgtgaatgtccgtttc	180
Db	121	CAAGAAGGTTTCAAAAACCTTCAGCGAGCAGCTCTGAAGAAATCCTGTGAATGTGCCGPTTC	180
Qy	181	ctctaaataccagacagtgtaaaaattacagcaatatattttttattccctctaatt	240
Db	181	CTCTAAATACACAGACTTGAAAATTTACAGCAATATATATTTTATTCCTCTAAAT	240
Qy	241	caagatacctacactggtttatatctctaaatgaattggctggaaactcctttatgatt	300
Db	241	CAAGGATACCTTACCTGGTTTATATCTTAAATGAATTTGGCTGGAACTCCTTTATGATAT	300
Qy	301	ctgcagacactgttaataataccagagaaacagctttgtactgtcgaaatcttgctttcac	360
Db	301	CTGCAGCACCTGTAAATAATACCCAGAGAACAGCTTTGTCTACTGCGAATCTTGGCTTCAC	360
Qy	361	tgcataccccctccatggacaaaagtcagagtaagcgctagatcccttaataagtt	420
Db	361	TGCCATCCCCCTCCATGGACAAATGAGTCAGAGTAAGCGCTAGGATCCCTTAAATAGTT	420
Qy	421	taagccaaagcccggttccattctcttagcaactgaagtgccagcgaggtttggacat	480
Db	421	TAAGCCAAAGCCCGTTCATCTCTTCTAGCAACTGACGTTGCCAGCGGAGTTTGGACAT	480
Qy	481	acctatgtagatgtggtgtcaactttgacattcctaccattccagaggtattacatcca	540
Db	481	ACCTCATGTAGATGTGGTTGTCAACTTTGACATTCCTACCCATTCCAAGGATTTACATCCA	540
Qy	541	tgaagtagtcgaacagctagagctggcgctccggaaaggctattactttgtcaacaca	600
Db	541	TGAGTAGTGTGAACAGCTAGAGCTGGCGCTCCGAAAGGCTATTACTTTGTTCACACA	600
Qy	601	gtataatgtgaactctccagcgatagaacacttaattgggaagaactaccaggttt	660
Db	601	GTATGATGTGGAACCTCTCCAGCGATAGACACTTAATTGGGAAGAACTACCAGGTTT	660
Qy	661	tccaacacaggtatgatgttatgactgacagaacgcgtcccccagcgatgtctccac	720
Db	661	TCCAACACAGGATGATGAGTTTATGATGCTGACAGAACGCGTCCCGCAGGATGTCTCCAC	720
Qy	721	caccgctgtcgaacccctgt	780
Db	721	CACCCTGTGTGAACCCCTGT	780
Qy	781	ccactatgtccgactcctcttcacagtcacacacctggagcagagccctgaacctac	840
Db	781	CCACTATGTCCGCATCTCTCTTCATCGAGTCCCACTTGAGCGCAGACCCCTGAACCTAC	840
Qy	841	tgaggggatggagagaacacagcagagctccccaagttggggggccccaatccctgggaca	900
Db	841	TGAGGGGATGGAGAGAACACAGCAGAGCTCCCAAGTTGGGGGCCCATTCCTCTGGGGACA	900
Qy	901	agccaatctctactcctctcgaactacagagatgtcagttattttgggaaattgggc	960
Db	901	AGCCCATCTCTGACTCTCTCGAACTACAGGGATGTGAGATTTTGGGAAATTTGGGC	960
Qy	961	tgggaacgctcccaaaaacttcaactgttgcctttgacactggctcctcccaatctctggg	1020
Db	961	TGGGAACGCTCCACAAAACCTTCACTGTTGCTTTTGACACTGGCTCCCTCCAATCTCTGGG	1020

Qy	1021	tccgtccaggagatgcacacttcttcagtgctccctgtcgtgtgtacaccacgattgac	1080
Db	1021	TCCGTCCAGGAGATGCCACTTCTTCAGTGTGCCCTGTGGTTACACCACCGATTGATC	1080
Qy	1081	ccaaagcctctagctctccagggccaaatgggaccaaagtgtccattcaatgaactg	1140
Db	1081	CCAAAGCCTCTAGCTCCTTCCAGGGCAATGGACCAAGTTTGCCTTCAATATGAACATG	1140
Qy	1141	ggcgggtagatgaatcctgagcgagagacaagctgactattgtggaatcaagggtgc	1200
Db	1141	GGCGGTAGATGAATCCTTAGCGAGGACAAGCTGACTATTTGGTGAATCAAGGTGAT	1200
Qy	1201	cagtgtatttccggggaggtctctctggagcccgagcctgtctctcgttttgcocatttg	1260
Db	1201	CAGTGATTTTGGGGAGGCTCTCTGGGACCCAGCCTGCTCTGCTTTTGGCCATTTTG	1260
Qy	1261	atgggatatgggctccgttttcccatctctgtctgtggaaggagttcgccccgatgg	1320
Db	1261	ATGGGATATTGGGCTTCGTTTCCCATTTCTGTGTGAAGGAGTTCGGCCCCCGATGG	1320
Qy	1321	atgtactgtgagcagggctatttgataaagcctgtctctctcttttaccctcaacaggg	1380
Db	1321	ATGTACTGTGAGCAGGGGCTATTGATTAAGCTGTCTCTCTTCTTACCTCAACAGGS	1380
Qy	1381	acctgaagagcctgatggaggagagctgtctgtgggggctcgtggaccgggaacataca	1440
Db	1381	ACCCTGAAGAGCCTGATGGAGGAGAGCTGTCTGTGGGGGCTCGGACCCGGCACATACA	1440
Qy	1441	tcccaacctcaactcgtgcagtcacggtccctcactactgcagatccacatgagc	1500
Db	1441	TCCCAACCTCTACCTTGTGTCAGTACAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1500
Qy	1501	gtgtgaaggtggggccagggctgactctctgtgccaagggtcgtgctgcactcctggata	1560
Db	1501	GTGTGAAGGTGGGCCAGGGCTGACTCTCTGTGCCAAGGCTGTGTGCTCATCTGGATA	1560
Qy	1561	cgggcagctccctcactcaacagaccactgagagatccgggcccctgcactgagccatg	1620
Db	1561	CGGGCAGCTCCCTCATACAGGACCCACTGAGGAGATCCGGGCCCTGCTATGACGCCATG	1620
Qy	1621	ggggaatcccccttgcctggctggggagtacatcatcctctgtcgtcggaaaatcccaagctcc	1680
Db	1621	GGGGAATCCCCCTTGTGCTGGGTGGGAGTACATCATCTGTCTCGAAATCCCAAGCTCC	1680
Qy	1681	ccgcagctcctcctcctctctctgggggggtctgggtttaaactcagggccccatgattagca	1740
Db	1681	CCGCAGTCTCCTTCTTCTTGGGGGGTCTGGTTTAACTTACCTCACGGGCCATGATACGTCA	1740
Qy	1741	tccagactactcaaaatggcgtccgctctgtgttccgggtttccagggccctggatgctc	1800
Db	1741	TCCAGACTACTCGAAATGGCGTCCGCCCTCTGTGTCCGGTTTCCAGGCCCTGGATGCTC	1800
Qy	1801	ctccgctcgcagggccctctctggtacccctcgtgacgtctctcttctggggacgtatgtggccg	1860
Db	1801	CTCCGCTGCAGGGCCCTTCTGGATCTCTGGATCTCTGTGTCCGGTTTCCAGGCCCTGGATGCTC	1860
Qy	1861	tcttcgaccggggagacatgaagaagcagcccggttgggacctggtcgccgctcgcactc	1920
Db	1861	TCTTCGACCGCGGGGACATGAAGAGCAGCCCGGGTGGGCGCTGGCGCGCTGCGACTC	1920
Qy	1921	gcggagcggaactcggatggggagagagactcgcagggcgaggttccccgggtgacgccccaa	1980
Db	1921	GCGGAGCGGACCTCGGATGGGAGAGACTGGCAGGCGCAGTTCCTCCCGGGTGGAGCCCAA	1980
Qy	1981	gtgaagcgcatgctcagcgggtgtcgcggaggttctcgtaccacagtaaaaaatccactat	2040
Db	1981	GTGAAGCGCATGCGCAGCGGGTGTCTCGGGAGGTCCTGCTACCCAGTAAAAATCCACTAT	2040
Qy	2041	ttccattgaaaaaataaaaaa 2060	
Db	2041	TTCCATTGAAAAAATAAAAAA 2060	

[illegible][illegible]









RESULT 7  
US-08-631-097-7  
; Sequence 7, Application US/08631097  
; Patent No. 5968816  
; GENERAL INFORMATION:  
; APPLICANT: Kimchi, Adi  
; TITLE OF INVENTION: Tumor Suppressor Genes,  
; TITLE OF INVENTION: Protein Enclosed Thereby, and Use of Said Genes and Protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wigman, Cohen, Leitner, & Myers, P.C.  
; STREET: 900 17th Street, N.W., Suite 1000  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/631,097  
; FILING DATE: 12-Apr-96  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/11598  
; FILING DATE: 12-Oct-94  
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Cohen, Herbert
; REGISTRATION NUMBER: 25,109
; REFERENCE/POCKET NUMBER: 0744.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)463-7700
; TELEFAX: (202)473-6915
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: No. 5968816 applicable
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; STRAIN: not applicable
; INDIVIDUAL ISOLATE: not applicable
; DEVELOPMENTAL STAGE: not applicable
; HAPLOTYPE: not applicable
; TISSUE TYPE: blood
; CELL TYPE: leucocyte
; CELL LINE: HeLa
; ORGANELLES: not applicable
; IMMEDIATE SOURCE:
; LIBRARY: not applicable
; CLONE: not applicable
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: not applicable
; MAP POSITION: not applicable
; UNITS: not applicable
; FEATURE:
; NAME/KEY: This is the DNA sequence
; LOCATION: claimed in 15(vi) as the Cathepsin gene in FIG. 15.
; IDENTIFICATION METHOD: experiment-
; OTHER INFORMATION: prevention of IFN-2
; OTHER INFORMATION: promoted cell death
; PUBLIC INFORMATION: not available
; US-08-631-097-7

Query Match          15.4%; Score 316.6; DB 2; Length 2038;
Best Local Similarity 57.9%; Pred. No. 6.4e-80;
Matches 602; Conservative 0; Mismatches 404; Indels 33; Gaps

QY 878 gggggggcccatccctctggggacaagcccatctctgtacctctctgaactacaggatgt 937
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 GCGGCTGCCAGCGGTGACCGAGGGGCCCATTCGCCAGGTGTCAGAACTACATGGAGCG 281

QY 938 gcagtatattgggaaattggctgggaagcctcccaaaacttcactgttgctttga 997
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 CCAGTACTACGGGGAGATGGCATCGGACGCCGCCCGCCAGTCTACAGTCGTCTTCGA 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 998 cactggctcctccaatctctgggtcccgctccaggagatgccactcttcagtgtgccctg 1057
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 342 CACGGGCTCTCCACCTGTGGGTCCCTCCATCCATCCATGCAAACTGCTGGACATCGCTTG 401
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1058 ctggtttacacaccgatttgatcccaagcctctagtctcttcaggccaatgggaccaa 1117
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 CTGGATPCCACCAAGTACAACAGGCGACAAGTCCAGCACCTTACCTGAAGATGGTACCTC 461
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1118 gtttgcattcaataatggaactggcgcggttagatggaatcctgacgcagacaagctga- 1176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 462 GTTTGATATCCACTATGTGCTCGGGCAGCCTCTCCGGGTACCTGACCCAGACACTGTGTC 521
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1177 -----ctattggtggaatcaagggtgtcatcagt 1204
;
; 522 GTGCGCCTGCCAGTCAGCGTCTGTAGCCCTCTGCGCCTGTCAAAGTGGAGAGGCA 581

```



QY 1205 gatttctggggagggctctctggagccagcctgggtctctgttctgttggccattttgatgg 1264  
DB 582 GGTCTTTGGGGAGGGCCACCAAGCAGCCAGGCGATCACCTTCATCGCAGCGAAGTTCGATGG 641  
QY 1265 gatatgggctcgtgttttccattctctgtgtggaaggagttcgccccccgatggatgt 1324  
DB 642 CATCTGGGCATGGCTACCCCGCGATCTCCGTCACAAACAGCTGCTGCGCGTCTTCGACAA 701  
QY 1325 actggtgagcagggctatggataagccctgtctctctctctctctctctctctctctctct 1384  
DB 702 CCTGATGACAGAAAGCTGGTGACCAAGCAATCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 761  
QY 1385 tgaagagcctgatggagagagctgtctctgggggctcgagcccgccacactatcccc 1444  
DB 762 AGATCGCGAGCCTGGGGGTGAGCTGATGCTGGTGGCAGACACTCCAAAGTATTACAAGGG 821  
QY 1445 accctcaccctctgtgacagtaeaggtccctctgctctctctctctctctctctctctctct 1504  
DB 822 TTCTCTGTCTACCTGAATGTACCCGCAAGGCTTACTGGCAGGTCCACCTGGACACAGGT 881  
QY 1505 gaaggtggccagggctgactctctgtgccaaggctgtgtgctgctgctgctgctgctgctg 1564  
DB 882 GGAGTGGCCAGCGGCTGACCTCTGTGAAGGAGGCGCTGTGAGGCCATTTGGACACAGG 941  
QY 1565 cagctccctcatcacagaccactgagagatccgggcccctgcatgagccattggggg 1624  
DB 942 CACTTCCCTCATGTGGGCGCGGTGATGAGTGGCGAGCTTGCAGAAAGGCCATCGGGGC 1001  
QY 1625 aatccctctgtggtgggagtaacatctctgtctgctggaatacccaagctcccgcc 1684  
DB 1002 CGTGGCGGTGATTCAGGGCGAGTACATGCCCTGTGAGAGGCTGTCCACCCCTGCCCGC 1061  
QY 1685 agtctccctctctctgtggggggtgtgttaacctcagcccgcccatgattacatccca 1744  
DB 1062 GATCACACTGAAGCTGGGAGGCAAGGCTACAAAGCTGTCCCCAGAGGACTACACGCTCAA 1121  
QY 1745 gactactgaaatggcgtccgctctgttccgggtttccaggccctggatgctccctcc 1804  
DB 1122 GGTGTGCGAGCGCGGAGACCCCTCTGCTGAGCGGCTTCATGGGCGATGGACATCCCGCC 1181  
QY 1805 gctcagcagggccctctctgagctcgtgagctctctctctctctctctctctctctctct 1864  
DB 1182 ACCCAGCGGCGACCTGTGATCTCTGGCGACGCTTTCATCGGCGCGCTACTACACTGTGTT 1241  
QY 1865 qcaccgcggggacatgaag 1883  
DB 1242 TGACCGTGACACACAGG 1260

## RESULT 8

US-08-810-712-11  
; Sequence 11, Application US/08810712G  
; Patent No. 6160106  
; GENERAL INFORMATION:  
; APPLICANT: Yeda Research and Development Co. LTD  
; TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
; FILE OF INVENTION: Use of said Genes and Proteins  
; FILE REFERENCE: sequence list  
; CURRENT APPLICATION NUMBER: US/08/810,712G  
; CURRENT FILING DATE: 1997-03-03  
; EARLIER APPLICATION NUMBER: PCT/US94/11598  
; EARLIER FILING DATE: 1994-10-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 2038  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-08-810-712-11  
Query Match 15.4%; Score 316.6; DB 4; Length 2038;  
Best Local Similarity 57.9%; Pred. No. 6.4e-80;

Matches 602; Conservative 0; Mismatches 404; Indels 33; Gaps 1;  
QY 878 gggggcccatccctcctgggagcaagcccatctctctctctctctctctctctctctctctct 937  
DB 222 ggcgtgtgcagccgtgacagaggggcccattccctcctcctcctcctcctcctcctcctcctc 281  
QY 938 cgagtattttggggaattgggctgggaacgctcccaaaaacttctcactgttgcctttga 997  
DB 282 ccagtactacagggagattggcatcggaacgccccccagtgcttccagtgctgttcga 341  
QY 998 cactggctcctcaatctctgggtcccgctccaggagagatgccaactcttctcagtgctccctg 1057  
DB 342 cagggctcctcaaatctgtgggtccctcctcctcctcctcctcctcctcctcctcctcctg 401  
QY 1058 ctggttccacacagatttgatcccaagcctcctagctccttccagggccaatggagaccaa 1117  
DB 402 ctgggtaccacaaagtaacacagcgaagcccgacacctctggaagaa tggtaacctc 461  
QY 1118 gtttgccttcaataggaactgggctgtagatggaatcctgagcgaggaacaagtga- 1176  
DB 462 gtttgacatccactatgctcgtggcagcctctccgggtacctgagccaggaacactgtgtc 521  
QY 1177 -----ctattggggaatacaagggtgcatcagt 1204  
DB 522 ggtgcccctgccaagtcagcgtctcagcctctgcccctggcggtgtcaaaagtggagagca 581  
QY 1205 gatttctgggaggtctctctggagcccgagcctgtctctctctctctctctctctctctctct 1264  
DB 582 ggtcctttggggagcccaagcagcagggcatcacttctcctcagcagcaagtctcgatgg 641  
QY 1265 gatattgggctcgttttcccatctctgtgtggaaggagttcgcccccgatggatgt 1324  
DB 642 cactcctgggcatggcctaccgcccatctccgtcaacaacgtgctgcccgtcttcgacaa 701  
QY 1325 actgtgtgagcaggggctattggaataagcctgtctctctctctctctctctctctctctctct 1384  
DB 702 cctgtatgagcaagaagctgtggaacagaaacatctctctctctcactgagcagggaccc 761  
QY 1385 tgaagagcctgtgagagagagctgtcctgggggctcgagcccgccacactatcccc 1444  
DB 762 agatcgcgacgctgtgggtgagctgtgctgggtggcagagctgagcagactccaagtattacaagg 821  
QY 1445 accctcaccctctgtgccaagtcacagctcctcctcctcctcctcctcctcctcctcctcctc 1504  
DB 822 ttctctgtcctacctgaatgtcacccgcaaggcctactgagcaggtccacctggaccaggt 881  
QY 1505 gaaggtgggcccagggctgactctctgtgccaaggctgtgtgctcctcctcctcctcctcctc 1564  
DB 882 ggaggtggccagcggtgacccctgtgcaaggagggctgtgagggcctgtgagggcctgtggaacagg 941  
QY 1565 cagctccctcatcacagaccacactgagagatccgggcccctgcatgagccattggggg 1624  
DB 942 cacttccctcatggtgggcccgtgtagatgagtgagtgagtgagtgagtgagtgagtgagtgag 1001  
QY 1625 aatccctctgtggtgggagtaacatctctgtctcctggaatacccaagctcccgcc 1684  
DB 1002 cgtgcccgtgattcagggcgagtagatgacccctgtgagaagggtgtccacctgcccgc 1061  
QY 1685 agtctcctcctctctgtggggggtgtgttaacctcagcggccccatgattacatccca 1744  
DB 1062 gatcacactgaagctgggggcaagggtacaaagctgtccccagaggaactcacgctcaa 1121  
QY 1745 gactactgaaatggcgtccgctctgttccgggtttccaggccctggatgctccctcc 1804  
DB 1122 ggtgtcagcagccgggaagaccctctgctgagcggtctcattggtgagcattgggaatcccgcc 1181  
QY 1805 gctcagcagggccctctctgagctcgtgacgtctctctctctctctctctctctctctctct 1864  
DB 1182 acccagcgggcccactctgactcgtggcgacgtctctctctctctctctctctctctctctct 1241  
QY 1865 cgacgcggggagacatgaag 1883  
DB 1242 tgaccgtgacaacaacagg 1260





Db	200	CCAGTTCCACGAGTCTGCTCAATGGACACAGAGTGCCAAAGNAACCCCTCATCAACTACTT	259
Qy	932	ggatgtagcagtatatttggggaaattggcgtgggaacgctcccaaaaacttcaactgttgc	991
Db	260	GGATATGGAAATACTTCGGCACTATCTCCATTGGCTCCCCACACACAGAACTTCACTGTGCAT	319
Qy	992	ctttgacactggcctccaactctctgggtccggtccagggagatgccaactcttcacgtgt	1051
Db	320	CTTCGACACTGGCTCTCCAACCTCTGGTGGCCCTCTGTG-----TACTGCACATAGCCC	373
Qy	1052	gcctgctggtttacacacccgattgatcccaaaagcctctagctccttccaggccaatgg	1111
Db	374	AGCCTGCAAGCGCAGCAGCAGGTTCACGCTTCCAGTCCAGACATACAGCCAGCCAGG	433
Qy	1112	gaccaagtgtgccattcaatatggaactggcgggtagatggaatcctgagcggaggcaa	1171
Db	434	TCAATCTTTCTCCATTACAGATATGGAACCGGAGAGCTTGTCCGGGATCATTTGGAGCGGACCA	493
Qy	1172	gctgactattggtggaatacaagggtgcatacagtgatttctggggaggctctctctggagcc	1231
Db	494	AGTCTCTGTGGGAAGGACTAAACCGTGTTGGCCAGCAGTTTGGAGAAAGTGTACAGAGACC	553
Qy	1232	cagcctggtcttcgcttttgccatttgcattgtagatattgggctcggttttcccatct	1291
Db	554	AGGCCAGACCTTTTGTGGATGTCAGAGATTTTGAATTTGGGCTTGGGCTACCCCTCCTT	613
Qy	1292	gtctggtgaaggagttcggcccccgatggatgtaactggtgagcaggggctattgataa	1351
Db	614	GGCTGTGGGAGGAGTGACTCCAGTATTTGACAAATGATGGCTCAGAACTGTGTGGACTT	673
Qy	1352	gcctgtcttctccttttacctcaacagggaacctgaagagcctgtagtggaggagcgtggt	1411
Db	674	GCGAGTGTTTCTGTCTACATGACAGTAAACCAGAAAGGTGGTCCGGGAGCGAGCTGAT	733
Qy	1412	cctggggggtctcgaccgcggcacactacatccccccctcactctcgtccagatcaggt	1471
Db	734	TTTTGGAGGCTACGACCACCTCCATTCTCTGTGGAGGCTGAAITGGTCCAGTACCAA	793
Qy	1472	ccctgcctactggcagatccacatgga	1498
Db	794	GCAAGCTTACTGGCAGATTGCACGTGA	820

RESULT 12

RESOL 12  
US-08-846-021A-6

; Sequence 6, Application US/08846021A

; Patent No. 5948682

; GENERAL INFORMATION:

APPLICANT: Moloney, Maurice M.

**TITLE OF INVENTION:** Preparation of Heterologous Proteins on

**TITLE OF INVENTION:** Oil Bodies

; NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: BERESKIN & PARR

STREET: 40 King Street West

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5H 3Y2

ELF: MCH 312  
: COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```

;
;
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0  Version #1.30

```

SOFTWARE: PatentIn Rel  
: CURRENT APPLICATION DATA:

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 00 045 031

```

;; APPLICATION NUMBER: US/08/87  
; FILING DATE: April 25, 1997

FILING DATE: April 25,  
CLASSIFICATION: 900

CLASSIFICATION: 800  
ATTORNEY / AGENT INFORMATION

ATTORNEY/AGENT INFORMATION:

NAME: Gravelle, Micheline

REGISTRATION NUMBER: 40,261

```
Db 2475 GAACAGTACGGTGATGTTGACATCGACTGCGACAACCTGAGCTACATGCCCACTGTGTGT 2534
Qy 1691 ctt 1693
Db 2535 CTT 2537

RESULT 13
5217891-14
;PATENT NO. 5217891
;APPLICANT: BRAKE, ANTHONY J.;VAN DEN BERG, JOHAN A.
;TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES
;A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
;POLYPEPTIDES
;NUMBER OF SEQUENCES: 23
;CURRENT APPLICATION DATA:
;APPLICATION NUMBER: US/08/507,398
;FILING DATE: 09-APR-1990
;PRIOR APPLICATION DATA:
;APPLICATION NUMBER: 78,551
;FILING DATE: 28-JUL-1987
;SEQ ID NO:14:
;LENGTH: 2732
5217891-14

Query Match      8.1%; Score 167.8; DB 6; Length 2732;
Best Local Similarity 52.0%; Pred. No. 1..le-37;
Matches 434; Conservative 0; Mismatches 387; Indels 14; Gaps 2;

Qy 911 cgtacctctctgaactacaggatgtgcagtattttggggaattggctgggaacgcc 970
Db 828 cgtccctctgacaaactacctggacagtcagtcactttgggaagatctacctcggaacccc 887
Qy 971 tccacaaactcaactgttgcctttgacactgctcctcccaatctctctgggtccctccag 1030
Db 888 gccacaggatgtccactgctgtttgacactggtcctctgactctgtgtggtacctctca- 946
Qy 1031 gagatgccactcttcagtgctccctgctgtgttacacacccagttgtatcccaaaccttc 1090
Db 947 -----tctactgcaagaaatgctgcaaaacacacagcgtctcgaccgcgagaaagtc 1001
Qy 1091 tagctctcttcaggccaaatggaccaaatgttgccattcaatatgaaactggcggttaga 1150
Db 1002 gtccacctccagaacctgggaagccctgtctctcactacacggagcagcagatgca 1061
Qy 1151 tgaatactgacggagcaactgactattgttggaatcaagggtgcatcagtgatttt 1210
Db 1062 gggcatctgggtctatgacacgtcactgtctccaaactgttggaatccacagacagtc 1121
Qy 1211 cggggaggctctctggagccagcctggtctctcttttgcctattttgatggatatt 1270
Db 1122 agcctgagcaccagagccggagcgtctctcactatgcccgaattcgacggagatcct 1181
Qy 1271 gggcctcgttttccattctgtctgtggaaggagttggtggcccccagatggatgtactggt 1330
Db 1182 ggggatggcctaccctcgtcgtcctcagagctactcgataccctgtgttgacaacatgat 1241
Qy 1331 ggagcaggggctattggataaacctctctctctcttttacctcaacacagggaacctgaaga 1390
Db 1242 gaacagggcacctggcgcccaagacctgtctcgttttacctatggacagg-----aat 1293
Qy 1391 gcctgatggagagagctggtcctggggggtcgcgaccgcgcacactacatcccacccct 1450
Db 1294 ggcceaaagagacatgctcaacgtcgtgggcccgaatcgaccgtcctactacaaagggtccct 1353
Qy 1451 caccctcgtgcagtcacaggtccctcgtcctactgctcagatccacatggagcgtgtgaaggt 1510
Db 1354 gcatgtgggtgcccgtgacagtcagcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1413
Qy 1511 gggccacagggtgactctctctgtgcacaaagggtgtgtcgtccatcctcgtgatacgggacgctc 1570
Db 1414 cagcgggtgtggtgtggcctgtgaggtgtggtgtcagggccatcctcgtgacacgggacacctc 1473
```

```
Qy 1571 cctcatcacagaccactgaggagatcgcgggcccctgcatgcagccattgggggaatccc 1630
Db 1474 caagctggtcgggcccagcagcagacatcctcaacatccagcagccattggagccacaca 1533
Qy 1631 cttgctggtcgggagtagatcatcctctgtctcggaatcccaagctcccgcagctctc 1690
Db 1534 gaaccagtagctgagtttgacatcgactgcgacaacctgagctacatgcccactgtggt 1593
Qy 1691 cttcctcttgggggggtctggttttaacctcaacggcccatgattacgtcatccag 1745
Db 1594 ctttgatcaaatggcaaatgtaccactgaccctcctcgcctatataccagccag 1648

RESULT 14
US-08-328-314-1
; Sequence 1, Application US/08328314
; Patent No. 5674728
; GENERAL INFORMATION:
; APPLICANT: Buxton, Frank
; APPLICANT: Jarai, Gabor
; APPLICANT: Visser, Jacob
; TITLE OF INVENTION: Fungal Protease
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,314
; FILING DATE: TBA
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprulli, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 4-19746/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8615
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2875 base pairs
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; TYPE: nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: pepe
; STRAIN: Aspergillus niger N400
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1269..1370, 1462..1612, 1669..2323, 2382
; LOCATION: ..2667)
; OTHER INFORMATION: /function= "Aspartic Protease"
; OTHER INFORMATION: /product= "PEPE"
; OTHER INFORMATION: /gene= "pepe"
; FEATURE:
; NAME/KEY: intron
; LOCATION: order(1371..1461, 1613..1668, 2324..2381)
; FEATURE:
; NAME/KEY: exon
; LOCATION: join(1269..1370, 1462..1612, 1669..2323, 2382
; LOCATION: ..2667)
```



```

Db 2023 AGATTGTTCTCCTCCCTTCTACAAACATGCTTGACAGGGGCTCCTCGACGAGCGGTCCTTG 2082
QY 1363 ccttttacctcaacaggaccctgaagacctgatggaggagagctgtccttgggggct 1422
Db 2083 CCTTCTACCTTGGAGACACCAACAAGGAGGGTGACGAGTCCGTGGGACCTTCGGTGGTG 2142
QY 1423 cggaccgggcacactacatccccaccctcaccttcgtgccagtcacggtccctgcctact 1482
Db 2143 TCGACAAGGACCCTACACCGGCGAGCTGATCAAGATCCCCCTCCGGCGCAAGGCTTACT 2202
QY 1483 ggcagatccacatggagcgtgtgaagggtggccagggtgactctctgtgccaaaggct 1542
Db 2203 GGGAGGTTGAGCTTGACGCCATTGCTCTTGGCGATGATGTTGCTGAGATGG---AGAAC 2259
QY 1543 gtgctgccatcctgatacgggcacgctccctcactcacaggaccactga 1591
Db 2260 CCGGTGTCATTCTGGACACTGCTACCTCCCTGATTGCTCTGCTGCTGA 2308

```

Search completed: November 17, 2001, 14:48:23  
Job time: 7181 sec

THIS PAGE BLANK (uspro)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 12:48:42 ; Search time 115.94 Seconds  
(without alignments)  
990.377 Million cell updates/sec

Title: US-09-700-770-1  
Perfect score: 507  
Sequence: 1 ggcaagtgaaccactggct.....aataaagcaatgaatacatt 507

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507	100.0	507	2	US-08-964-725-4
2	499.4	98.5	519	2	US-08-964-725-5
3	263	51.9	263	2	US-08-964-725-2
4	263	51.9	269	2	US-08-964-725-1
5	225	44.4	225	2	US-08-964-725-3
6	38	7.5	7218	1	US-08-232-463-14
7	34.8	6.9	7218	1	US-08-232-463-14
8	32.6	6.4	3471	2	US-08-715-568A-2
9	31.8	6.3	80246	4	US-09-078-294-4
10	31.8	6.3	80595	4	US-09-078-294-3
11	31.6	6.2	978	3	US-09-267-031-13
12	30	5.9	5962	6	5386025-5
13	30	5.9	5975	1	US-08-404-354B-1
14	30	5.9	5975	1	US-08-314-083B-1
15	30	5.9	5975	1	US-08-435-675B-1
16	30	5.9	5975	3	US-08-336-257A-3
17	30	5.9	5975	3	US-08-884-599-1
18	29.4	5.8	2338	1	US-08-425-069-1
19	29.4	5.8	2338	2	US-08-317-844B-1
20	29.4	5.8	3066	4	US-09-086-912-1
21	29	5.7	289	4	US-09-007-005-17
22	29	5.7	289	4	US-09-244-796-17
23	29	5.7	72928	3	US-09-009-913-1
24	28.6	5.6	87350	3	US-08-781-891-79
25	28.4	5.6	545	4	US-09-073-297-17
26	28.4	5.6	3489	2	US-08-728-323A-1
27	28.4	5.6	4181	1	US-07-670-611-1

Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 20, Appli  
Sequence 20, Appli  
Sequence 68, Appli  
Sequence 68, Appli  
Sequence 68, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 6, Appli  
Sequence 25, Appli  
Sequence 11, Appli  
Sequence 11, Appli  
Sequence 11, Appli  
Sequence 11, Appli  
Sequence 64, Appli

US-08-220-674-1  
US-08-445-186-1  
US-08-446-549-1  
US-08-446-550-1  
US-08-770-379-20  
US-08-757-669A-20  
US-08-474-379C-68  
US-09-146-249A-68  
US-08-206-188B-68  
US-08-464-339A-1  
PCT-US94-14388-1  
US-07-861-458C-6  
US-08-209-521-25  
US-07-688-352C-11  
US-08-474-379C-11  
US-09-146-249A-11  
US-08-206-188B-11  
US-08-474-379C-64

ALIGNMENTS

RESULT 1  
US-08-964-725-4  
Sequence 4, Application US/08964725  
Patent No. 5939265  
GENERAL INFORMATION:

APPLICANT: COHEN, Maurice  
APPLICANT: FRIEDMAN, Paula N.  
APPLICANT: GORDON, Julian  
APPLICANT: HODGES, Steven C.  
APPLICANT: KLASS, Michael R.  
APPLICANT: KRATOCHVIL, Jon D.  
APPLICANT: ROBERTS-RAPP, Lisa  
APPLICANT: RUSSELL, John C.  
APPLICANT: STROUPE, Steven D.

TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
FOR DETECTING DISEASES OF THE LUNG  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA

ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA: US/08/964,725

APPLICATION NUMBER: 5386025

FILING DATE: 1997/09/17

PRIOR APPLICATION DATA:

CLASSIFICATION: 435

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Becker, Cheryl L.

REGISTRATION NUMBER: 35,441

REFERENCE/DOCKET NUMBER: 5997.US.P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847/935-1729

TELEX: 847/938-2623

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 507 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

5939265

US-08-964-725-4

Query Match 100.0%; Score 507; DB 2; Length 507;  
Best Local Similarity 100.0%; Pred. No. 2.6e-147;  
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagtggaaccactggtggtggttttctgtagatttttctgatttttaaaactctg 60  
Db 1 ggcagtggaaccactggtggtggttttctgtagatttttctgatttttaaaactctg 60

QY 61 aaaaatccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120  
Db 61 aaaaatccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120

QY 121 cctttagttactctgctactgctctctctcatcaaaaagtgcccttctggtgacaa 180  
Db 121 cctttagttactctgctactgctctctctcatcaaaaagtgcccttctggtgacaa 180

QY 181 gtggaaccttaactctggaacaacttctccctttatggtatccatcaaaagtctct 240  
Db 181 gtggaaccttaactctggaacaacttctccctttatggtatccatcaaaagtctct 240

QY 241 gaaaactctggcatttctgtgagcactgtgaggggctgaagagtggtataatga 300  
Db 241 gaaaactctggcatttctgtgagcactgtgaggggctgaagagtggtataatga 300

QY 301 gctgggaccagagcttctgaaactgtgaagaactgctgagggcgctatcacacttgg 360  
Db 301 gctgggaccagagcttctgaaactgtgaagaactgctgagggcgctatcacacttgg 360

QY 361 gtaecatcaagataaagagcgaggtggtgaggtgaggtgaggtgaggtgaggtgag 420  
Db 361 gtaecatcaagataaagagcgaggtggtgaggtgaggtgaggtgaggtgaggtgag 420

QY 421 tgcctgaacctgtctacaaattatagatcaaatgcccctaaatgtagtgaccctgaa 480  
Db 421 tgcctgaacctgtctacaaattatagatcaaatgcccctaaatgtagtgaccctgaa 480

QY 481 aagacaataaagcaatgaatacatt 507  
Db 481 aagacaataaagcaatgaatacatt 507

## RESULT 2

US-08-964-725-5  
Sequence 5, Application US/08964725  
Patent No. 5939265

## GENERAL INFORMATION:

APPLICANT: COHEN, Maurice  
APPLICANT: FRIEDMAN, Paula N.  
APPLICANT: GORDON, Julian  
APPLICANT: HODGES, Steven C.  
APPLICANT: KLASS, Michael R.  
APPLICANT: KRATOCHVIL, Jon D.  
APPLICANT: ROBERTS-RAPP, Lisa  
APPLICANT: RUSSELL, John C.  
APPLICANT: STROUPE, Steven D.

TITLE OF INVENTION: REAGENTS AND METHODS USEFUL

FOR DETECTING DISEASES OF THE LUNG

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories

STREET: 100 Abbott Park Road

CITY: Abbott Park

STATE: IL

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/964,725  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 5997.US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 519 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-964-725-5

Query Match 98.5%; Score 499.4; DB 2; Length 519;  
Best Local Similarity 99.8%; Pred. No. 5.8e-145;  
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 gtggaaccaactggctgggtggttttctgtagatttttctgatttttaaaactctgaaaaa 65  
Db 19 GTGGAACCACTGGCTGGTGGATTTTGTAGATTTTCTGATTTTAAACTCTCTGAAAAA 78

QY 66 tatccagataactgtcatgaagctggttaactatcttctgctggtgaccatcagccttt 125  
Db 79 TATCCAGATAACTGTCAAGCTGGTAACATATCTTCTGCTGGTGACCATCAGCCTTT 138

QY 126 gtagttactctgctactgctctctctctcaacaagaagtcaccttctggtgacaagtgg 185  
Db 139 GTAGTTACTCTGCTACTGCTTCTCTCATCAACAAAGTGCCCTTCTGTTGACAAAGTTGG 198

QY 186 cactttactctggacaacattcttcttctgattgattcattgattcatttcttctgaaaa 245  
Db 199 CACTTTACTCTGGACACATCTTCCCTTTATGATCCATTAAGCTTCTTCTGAAAAA 258

QY 246 ctctgggcatcttctgtgagcacttgtgaggggctgaggaagtggttaaatgagctgg 305  
Db 259 CTCTGGCATTTCTGTGAGCACCTTGTGGAGGGCTAAGGAAGTGTGTAAATGAGCTGG 318

QY 306 gaccagagcttctgaaactgtgaagaactctgagggcgctatcacacttgggtgac 365  
Db 319 GACCAGAGCTTCTGAAGCTGTGAAGAACTGCTGGAGGCGCTATCACACTTGGTGTGAC 378

QY 366 atcaagataaagagcgagggtggtgaggggctgaggaagtgagcctctatccctccctgct 425  
Db 379 ATCAAGATAAAGAGCGAGGTGGATGGGATGGAAGATGATGCTCTATCTCTCCCTGCT 438

QY 426 gaaacctgttctaccaattatagatcaaatgccctaaatgtagtgacctgtaaaagga 485  
Db 439 GAAACCTGTTTACCAATTATAGATCAAAATGCCCTTAAATGTAGTGACCCCTGAAAAAGGA 498

QY 486 caataaagcaatgaatacat 506  
Db 499 CAATAAAGCAATGAATACCT 519

## RESULT 3

US-08-964-725-2  
Sequence 2, Application US/08964725  
Patent No. 5939265  
GENERAL INFORMATION:  
APPLICANT: COHEN, Maurice  
APPLICANT: FRIEDMAN, Paula N.  
APPLICANT: GORDON, Julian

APPLICANT: HODGES, Steven C.  
APPLICANT: KLASS, Michael R.  
APPLICANT: KRATOCHVIL, Jon D.  
APPLICANT: ROBERTS-RAPP, Lisa  
APPLICANT: RUSSELL, John C.  
APPLICANT: STROUPE, Steven D.  
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
FOR DETECTING DISEASES OF THE LUNG  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/964,725  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 5997.US.P1  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-964-725-2

Query Match 51.9% Score 263; DB 2; Length 263;  
Best Local Similarity 100.0%; Pred. No. 4.1e-72;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 tgctactgccttcctcatcaacaaagtgccttcctctgttgacaaagtggcacctttacc 195  
|||||  
Db 1 TGCTACTGCCTTCCTCATCAACAAAGTGCCCTTCCTCTGTGACAAAGTTGGCACCTTTACC 60  
QY 196 tctggacaaacatttccctttatggatccattaaagcttcttgaaaaactctgggcat 255  
|||||  
Db 61 TCTGGACAACATTTCTCCCTTTATGSGATCCATTAAAGCTTCTTGAAAACTCTGGGCAT 120  
QY 256 tctgttgagacactgtgagggctaaagaaagtgttaaatgagctggacacagagcc 315  
|||||  
Db 121 TTCGTGTGACACCTTGTGGAGGGGCTTAAGGAAGTGTGAATGAGCTGGACAGAGGC 180  
QY 316 tcttgaagctgtgaagaaactgtggaggcgctatcacacttggtgtgacatcaagataa 375  
|||||  
Db 181 TTCTGAAGCTGTGAAGAACTGCTGGAGGCGCTATCACACTTGTGTGACATCAAGATAA 240  
QY 376 agagcgaggtgagtgaggatgg 398  
|||||  
Db 241 AGACGGAGGTGGATGGGATGG 263

RESULT 4  
US-08-964-725-1  
; Sequence 1, Application US/08964725

Patent No. 5939265  
GENERAL INFORMATION:  
APPLICANT: COHEN, Maurice  
APPLICANT: FRIEDMAN, Paula N.  
APPLICANT: GORDON, Julian  
APPLICANT: HODGES, Steven C.  
APPLICANT: KLASS, Michael R.  
APPLICANT: KRATOCHVIL, Jon D.  
APPLICANT: ROBERTS-RAPP, Lisa  
APPLICANT: RUSSELL, John C.  
APPLICANT: STROUPE, Steven D.  
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
FOR DETECTING DISEASES OF THE LUNG  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/964,725  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 5997.US.P1  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 269 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-964-725-1

Query Match 51.9% Score 263; DB 2; Length 269;  
Best Local Similarity 100.0%; Pred. No. 4.2e-72;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gtggaacacactggctgtggtgatttgcctagattttctctgatttttaaacctcctgaaaaa 65  
|||||  
Db 7 GTGGAACCACTGGCTTGGTGATTTTGTAGATTTTCTGATTTTAAACTCCTCGAAAAA 66  
QY 66 tatcccgataactgtgctgaagctggttaactatctctctgctggtgacacatcagccttt 125  
|||||  
Db 67 TATCCAGATAAATGTGATGAAGCTGTAATCTCTCTGCTGCTGACCATCAGCCTTT 126  
QY 126 gtagttactctgctactgccttctctcatcaacaaagtgcccttctctgttgacaaagtgg 185  
|||||  
Db 127 GTAGTTACTCTGCTACTGCTCTCTCTCATCAAAAGTGCCCTTCTCTGTTGACAAAGTTGG 186  
QY 186 cacctttacctctggacacacattcttccctttatggatccattaaagcttctctgaaaa 245  
|||||  
Db 187 CACCTTTACCTCTGGACAACATTTCTCCCTTTATGGATCCATTAAAGCTTCTTCTGAAAA 246  
QY 246 cctcggcgattctctgtgagcac 268  
|||||  
Db 247 CTCTGGGCATTTCTGTGAGCAC 269

Db 181 AATGTAGTACCCGCGTGAAGGACAAATAAAGCAATGAATACATT 225

RESULT 6  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
; US-08-232-463-14

Query Match 7.5%; Score 38; DB 1; Length 7218;  
Best Local Similarity 4.3%; Pred. No. 0.061;  
Matches 11; Conservative 144; Mismatches 99; Indels 0; Gaps 0;

Qy 18 gcttggtgatttctgtagattttctgatttttaaaactcctgaaataatccagataa 77  
Db 1048 GGTGCGAGGAGCTTGCATYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1107  
Qy 78 ctgcatgaagctggttaactatctctctggtgacatcagcctttagttactctg 137  
Db 1108 YVY 1167  
Qy 138 ctactgctctccatcaacaagaagtgccctctctctgttgacaaagtggacacttaccctc 197  
Db 1168 YVY 1227  
Qy 198 tgacacaactctctccctttatggtatccattaaagctctctctgaaactctggcat 257  
Db 1228 YVY 1287

Db 283 aaggaagtgtgtaaatgagctggaccagagcttctgaagctgtgaagaaactgctgga 342  
Db 1 AAGGAAGTGTGAATAGCTGGACCGAGGCTTCTGAAGCTGTGAAGAAACTGCTGGA 60  
Qy 343 ggcgtatcacacttggtgtgacatcaagataaagcgagtggtggtggatggaaga 402  
Db 61 GCGCTATCACACTTGGTGTGACATCAAGATAAAGCGGAGGTGGATGGGATGGAAGA 120  
Qy 403 tgatgctctatcctcctgctgaaactgttctaccaattatagatcaaatgacctaa 462  
Db 121 TGATGCTCTATCTCCCTCGCTGAACCTGTCTACCAATTATAGATCAAAATGCCCTAA 180  
Qy 463 aatgtagtaccctgaaagacaaaataaagaacatgaatcatt 507  
|||||

Query Match 44.4%; Score 225; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 2.1e-60;  
Matches .225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-08-964-725-3  
; Sequence 3, Application US/08964725  
; Patent No. 5939265  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, Maurice  
; APPLICANT: FRIEDMAN, Paula N.  
; APPLICANT: GORDON, Julian  
; APPLICANT: HODGES, Steven C.  
; APPLICANT: KLASS, Michael R.  
; APPLICANT: KRATOCHVIL, Jon D.  
; APPLICANT: ROBERTS-RAPP, Lisa  
; APPLICANT: RUSSELL, John C.  
; APPLICANT: STROUPE, Steven D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; IN DETECTING DISEASES OF THE LUNG  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/964,725  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 5997.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-964-725-3

```
Qy 258 ctgttgagcacctt 271
: : : : :
Db 1288 YYYYYYYYYYYY 1301

RESULT 7
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08/232.463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935.313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: ptzgpt-Fls
US-08-232-463-14

Query Match 6.9%; Score 34.8; DB 1; Length 7218;
Best Local Similarity 6.1%; Pred. NO. 0.6;
Matches 9; Conservative 91; Mismatches 48; Indels 0; Gaps 0;

Qy 274 ggaaggggctaaggagtgtgtaaatgagctgggaccagagccttctgaagctgtgaagaa 333
: : : : :
Db 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141

Qy 334 actgctgagcgctatcacacttggtgtgacatcaagataaagcgaggatggatggg 393
: : : : :
Db 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081

Qy 394 gatggaagatgatgtctctatctccct 421
: : : : :
Db 1080 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1053

RESULT 8

US-08-715-568A-2/c
: Sequence 2, Application US/08715568A
: Patent No. 5856463
: GENERAL INFORMATION:
: APPLICANT: Prydz, Hans Peter Blankenborg
: APPLICANT: Brede, Gaute
: TITLE OF INVENTION: PSKH-1 Ribozymes and Uses in Disease
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
: STREET: 600 South Avenue West
: CITY: Westfield
: STATE: NJ
: COUNTRY: USA
: ZIP: 07090-1497
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08/715.568A
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Foley, Shawn P.
: REFERENCE/DOCKET NUMBER: FORSK 3.0-002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 908-654-5000
: TELEFAX: 908-654-7866
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3471 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
US-08-715-568A-2

Query Match 6.4%; Score 32.6; DB 2; Length 3471;
Best Local Similarity 51.7%; Pred. No. 2;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 282 taagaaagtgtgtaaatgagctggagaccagagccttctgaagctgtgaagaaactgctgg 341
: : : : :
Db 1298 TATTGGAGCGTGTGGAGCGGCTGGAACCGGTGGACTGGGCAGATTGTGCTCTGGCAGC 1239

Qy 342 aggcgtatcacacttggtgtgacatcaagataaagcgaggatggatgggagtggaag 401
: : : : :
Db 1238 GCGAGGAGGCACGTTTAAAGAGAGTTCTGGGATATGGACGCGTGCAGGTTCTTCATGGATG 1179

Qy 402 atgatgctctatctctccctgcc 424
: : : : :
Db 1178 AAGAGGAGCCATGCTCACCACC 1156

RESULT 9
US-09-078-294-4
: Sequence 4, Application US/09078294
: Patent No. 6265211
: GENERAL INFORMATION:
: APPLICANT: Choo, Kong-Hong Andy
: APPLICANT: Du Sart, Desiree
: APPLICANT: Cancilla, Michael R.
: TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
: FILE REFERENCE: Davies Col
: CURRENT APPLICATION NUMBER: US/09/078.294
: CURRENT FILING DATE: 1998-05-13
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 80246
```



```

; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,354B
; FILING DATE: 15-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,083
; FILING DATE: 28-SEP-1994
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 08-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-53192
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 79...5700
; OTHER INFORMATION:
; US-08-404-354B-1

```

```

Query Match 5.9%; Score 30; DB 1; Length 5975;
Best Local Similarity 55.9%; Pred. No. 17;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 272 gtgaggggctaagaagtgtgtaaatgagctggagcagaggtcttgaagctgtgaag 331
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2979 GTGAGGGGGCTACTACTATGTGTACAGGACGGGACCCCGACGAGATGGAGCTGCGCCC 3038

QY 332 aaactgtggaggcgctatcacacttgggtgacatcaagat 373
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3039 CCGCAGTGGGATACAAATGACTTCCACTTTTGACAACTGTCT 3080

```

```

RESULT 14
US-08-314-083B-1
; Sequence 1, Application US/08314083B
; Patent No. 5686241
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/314,083B
; FILING DATE: 28-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 08-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-53191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 79...5700
; OTHER INFORMATION:
; US-08-314-083B-1

```

```

Query Match 5.9%; Score 30; DB 1; Length 5975;
Best Local Similarity 55.9%; Pred. No. 17;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 272 gtgaggggctaagaagtgtgtaaatgagctggagcagaggtcttgaagctgtgaag 331
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2979 GTGAGGGGGCTACTACTATGTGTACAGGACGGGACCCCGACGAGATGGAGCTGCGCCC 3038

QY 332 aaactgtggaggcgctatcacacttgggtgacatcaagat 373
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3039 CCGCAGTGGGATACAAATGACTTCCACTTTTGACAACTGTCT 3080

```

```

RESULT 15
US-08-435-675B-1
; Sequence 1, Application US/08435675B
; Patent No. 5710250
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

```
;; SOFTWARE: FastSeq Version 1.5
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/435,675B
;; FILING DATE: 05-MAY-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/314,083
;; FILING DATE: 28-SEP-1994
;; APPLICATION NUMBER: US 07/914,231
;; FILING DATE: 13-JUL-1992
;; APPLICATION NUMBER: US 07/603,751
;; FILING DATE: 08-NOV-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie L
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 6362-53193
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-238-0999
;; TELEFAX: 619-238-0062
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5975 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; * NAME/KEY: Coding Sequence
;; LOCATION: 79...5700
;; OTHER INFORMATION:
;; US-08-435-675B-1
```

```
Query Match          5.9%; Score 30; DB 1; Length 5975;
Best Local Similarity 55.9%; Pred. No. 17;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 272 gtggaggggcctaaggaagtgtgtaaatgagctgggaccagagccttctggaagctgtgaag 331
    ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 2979 GTGCAGGGGCTACTATGTGTACAGGCGGGGACCCACCGAGATGGAGCTGGCC 3038

Qy 332 aaactgtcggaggcgctatcacacttgggtgacatcaagat 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3039 CCGCCAGTGGATACAAATGACTTCCACTTTGACACACGTGCT 3080
```

Search completed: November 17, 2001, 14:47:27  
Job time: 7125 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 12:37:21 ; Search time 2570.26 seconds  
(without alignments)  
1864.635 Million cell updates/sec

Title: US-09-700-770-1  
Perfect score: 507

Sequence: 1 ggcaagtggaaacctggct.....aataagcaatgaatacatt 507

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	gb_est1.*
2:	gb_est2.*
3:	gb_est3.*
4:	gb_est4.*
5:	gb_est5.*
6:	gb_est6.*
7:	gb_est7.*
8:	gb_est8.*
9:	gb_est9.*
10:	gb_est10.*
11:	gb_est11.*
12:	gb_est12.*
13:	gb_est13.*
14:	gb_est14.*
15:	gb_est15.*
16:	gb_est16.*
17:	gb_est17.*
18:	gb_est18.*
19:	gb_est19.*
20:	gb_est20.*
21:	gb_est21.*
22:	gb_est22.*
23:	gb_est23.*
24:	gb_est24.*
25:	gb_est33.*
26:	gb_est34.*
27:	gb_est35.*
28:	gb_est36.*
29:	gb_est37.*
30:	gb_est38.*
31:	gb_est39.*
32:	gb_est40.*
33:	em_estba.*
34:	em_estfun.*
35:	em_esthum1.*
36:	em_esthum2.*
37:	em_esthum3.*
38:	em_esthum4.*
39:	em_esthum5.*
40:	em_esthum6.*
41:	em_esthum7.*
42:	em_esthum8.*
43:	em_esthum9.*
44:	em_esthum10.*
45:	em_esthum11.*
46:	em_esthum12.*
47:	em_esthum13.*
48:	em_esthum14.*
49:	em_esthum15.*
50:	em_esthum16.*
51:	em_esthum17.*
52:	em_esthum18.*
53:	em_esthum19.*
54:	em_esthum20.*
55:	em_esthum21.*
56:	em_esthum22.*
57:	em_esthum23.*
58:	em_esthum24.*
59:	em_esthum25.*
60:	em_esthum26.*
61:	em_esthum27.*
62:	em_esthum28.*
63:	em_estin1.*
64:	em_estin2.*
65:	em_estin3.*
66:	em_estin4.*
67:	em_estin5.*
68:	em_estom1.*
69:	em_estom2.*
70:	em_estov1.*
71:	em_estov2.*
72:	em_estpl1.*
73:	em_estpl2.*
74:	em_estpl3.*
75:	em_estpl4.*
76:	em_estpl5.*
77:	em_estpl6.*
78:	em_estpl7.*
79:	em_estpl8.*
80:	em_estpl9.*
81:	em_estpl10.*
82:	em_estro1.*
83:	em_estro2.*
84:	em_estro3.*
85:	em_estro4.*
86:	em_estro5.*
87:	em_estro6.*
88:	em_estro7.*
89:	em_estro8.*
90:	em_estro9.*
91:	em_estro10.*
92:	em_estro11.*
93:	em_estro12.*
94:	em_estro13.*
95:	em_estro14.*
96:	em_estro15.*
97:	em_estro16.*
98:	em_estro17.*
99:	em_estro18.*
100:	em_estro19.*
101:	em_estro20.*
102:	gb_est25.*
103:	gb_est26.*
104:	gb_est27.*
105:	gb_est28.*
106:	gb_est29.*
107:	gb_est30.*
108:	gb_est31.*
109:	gb_est32.*
110:	gb_est41.*
111:	gb_est42.*
112:	gb_est43.*
113:	gb_est44.*
114:	gb_est45.*
115:	gb_est46.*
116:	gb_est47.*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
C 1	505	99.6	516	148	BF433853 7q58h12.x
C 2	503.4	99.3	549	155	BG537963 602563644
C 3	503	99.2	529	155	BG545709 602572977
C 4	494.4	97.5	509	137	BG550203 7b50b11.x
C 5	493	97.2	433	19	AI355302 qt57e07.x
C 6	490.4	96.7	823	155	BG540454 602568773
C 7	482	95.1	757	155	BG541018 602570177
C 8	479.2	94.5	825	154	BG484055 602504659
C 9	462.8	91.3	474	17	AI193988 qe81c04.x
C 10	458	90.3	461	144	BF111035 7n31d06.x
C 11	456.8	90.1	462	166	BE348591 ht72d05.x
C 12	452	89.2	817	155	BG547781 602575659
C 13	449.8	88.7	510	187	N92483 zbl2g04.s1
C 14	446	88.0	464	14	AA992209 ct80f12.s
C 15	445	87.8	445	17	AI193016 qe9a09.x
C 16	444	87.6	450	19	AI355612 qt58b02.x
C 17	443	87.4	446	23	AI695085 we46f11.x
C 18	441	87.0	452	24	AI734244 zbl2g04.y
C 19	440	86.8	453	23	AI668642 zbl2g04.x
C 20	438	86.4	448	166	BE328852 hv97g12.x
C 21	434	85.6	435	18	AI263451 q107a12.x
C 22	433.2	85.4	438	18	AI299210 g03d02.x
C 23	428	84.4	439	148	BF433537 7q54g01.x
C 24	424.4	83.7	702	154	BG507111 601861755
C 25	422	83.2	422	18	AI298558 gn02d04.x
C 26	416	82.1	416	14	AA988921 or81f01.s
C 27	399	78.7	405	19	AI354759 qt55d11.x
C 28	396	78.1	398	14	AA988559 or84f04.s
C 29	394	77.7	418	190	W40141 z074b09.r1
C 30	369.4	72.9	567	137	BE552278 hy05d09.x
C 31	360	71.0	362	164	BE222065 hu05d04.x
C 32	355.4	70.1	359	23	AI693419 wd90g12.x
C 33	352.4	69.5	356	166	BE326306 hv99c02.x
C 34	345.4	68.1	360	148	BF433805 nab58e04.
C 35	337.6	66.6	348	170	BF826945 CM4-HN002
C 36	321.4	63.4	733	154	BG482860 602502458
C 37	319	62.9	334	116	AW449467 UI-H-B13-
C 38	316.4	62.4	387	190	W17168 zbl2g04.r1
C 39	299	59.0	311	18	AI265819 qx69a03.x
C 40	297.2	58.6	320	170	BF826946 CM4-HN002
C 41	295	58.2	310	156	D30977 HUML11966.H
C 42	286.4	56.5	290	156	D31546 HUML11712.Hu
C 43	282.6	55.7	333	156	D30961 HUML11877.H
C 44	281.4	55.5	299	156	D30979 HUML11982.H
C 45	279.6	55.1	291	156	D31534 HUML11612.Hu

## ALIGNMENTS

RESULT 1  
BF433853/c 516 bp mRNA EST 30-MAR-2001  
LOCUS 7q58h12.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3702767 3',  
DEFINITION mRNA sequence.  
ACCESSION BF433853  
VERSION BF433853.1 GI:11446183  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 516)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov

Seq primer: -40UP from Gibco  
High quality sequence stop: 481.  
Location/Qualifiers

## FEATURES

source  
1..516  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3702767"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI\_CGAP\_Lu5 was prepared, and ss circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonalido."  
BASE COUNT 142 a 121 c 112 g 141 t  
ORIGIN

Query Match 99.68; Score 505; DB 148; Length 516;  
Best Local Similarity 100.0%; Pred. NO. 3.6e-141;  
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcagtggaaccactggcttggtgatttgcagattttctgattttctaaactcctg 60  
Db 507 GGCAAGTGGACCACTGGCTGGTGGATTGCTAGATTTTCTGATTTTAACTCTG 448  
Qy 61 aaaaataccagataactgtcatgaagctggtgaactatctctctggtgaccatcag 120  
Db 447 AAAAATATCCAGATAAATGTGATGAAGCTGGTAACTATCTCTGCTGGTGCACATCAG 388  
Qy 121 cctttgtagttactctactgctctcctcatcaaaaagtgcccttctgttgacaa 180  
Db 387 CCTTTGTAGTTACTCTGCTACTGCTTCTCCTCATCAACAAAGTGCCCTTCTGTGACAA 328  
Qy 181 gttggcacccttacctctggacaacattcttccctttatggtatccattaaagctcttct 240  
Db 327 GTTGGCACCTTACCTCTGGACACATCTTCCCTTTATGGATCCATTAAGCTTCTTCT 268  
Qy 241 gaaaactctgggacatttctgttgagcaacctgtgaggggctgaaggaagtgttaaatga 300  
Db 267 GAAAACTCTGGCATTTCTGTTGAGCACCTTGTGGAGGGGCTAAAGGAAGTGTGAAATGA 208  
Qy 301 gctgggaccagagcttctctgaagctgtgaagaaactgctggaggcgctatcacacttggt 360  
Db 207 GCTGGGACCAAGAGGCTTCTTGAAGCTGTGAAGAAACTGCTGGAGGGCGCTATCACACTTGGT 148  
Qy 361 gtgacatcaagataaaagcggaggtggtggtggtggtggtggtggtggtggtggtggtggt 420  
Db 147 GTGACATCAAGTAAAGACGGAGGTGGATGGGATGGAGATGATGCTCTCTATCTCTCC 88  
Qy 421 tgcctgaacctgttctaccatattatagatcaaatgccctaaatagttagtgacctgtgaa 480  
Db 87 TGCCTGAAACCTGTTCTACCAATATATAGTCAAAATGCCCTAAATGTAGTAACTGACCGGTGAA 28  
Qy 481 aaggacaaaataaagcaaatgaatataca 505  
Db 27 AAGGACAAAATAAGCAATGAATACA 3

```

RESULT 2
BG537963 549 bp mRNA EST 03-APR-2001
LOCUS 60256364F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4688315 5',
DEFINITION mRNA sequence.
ACCESSION BG537963
VERSION BG537963.1 GI:13530195
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 549)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1502 row: h column: 12
High quality sequence stop: 515.
FEATURES
source
Location/Qualifiers
1..549
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4688315"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggccattggcc); Site_2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCCGACATG-dr(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 162 a 121 c 119 g 147 t
ORIGIN
Query Match 99.3%; Score 503.4; DB 155; Length 549;
Best Local Similarity 99.8%; Pred. No. 1.1e-140;
Matches 504; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gccaagtgaaccactgcttggtgatttgcctgattttctgatttttaaacctctg 60
Db 14 GGCAGTGGAACCACTGGCTTGGTGATTTCCTAGATTTTCTGATTTTAAACTCTGT 73
Qy 61 aaaaatccccagataactgtcatgaagctggttaactatcttctgctggtgaaccatcag 120
Db 74 AAAAAATATCCAGATAAATGTCATGAAGCTGTAATCTATCTTCCTGCTGGTGACCATCAG 133
Qy 121 ccttctgattactctgactgcttctctctatcaacaagaagtcctctctgttgacaa 180
Db 134 CCTTTGATTACTCTGCTACTGCTTCTCTCATCAACAAGTGCCCTTCTCTGTGACAA 193
Qy 181 gttggcaccctttacctcttgacaaacattcttcttcttgatgacattaaagcttctct 240
Db 194 GTTGGACCTTTACCTCTGGACAACTTCTCCCTTTATGATTCATTAAGCTTCTTCT 253
Qy 241 gaaactctggcattctctgttgagcacctgtgtgagggggttaaggaagtgtgtaataga 300
Db 254 GAAAACTCTGGGCATTTCTGTTGAGCACTTCTGTGGAGGGGCTAAGGAAGTCTGTAATGA 313

```

---

```

Qy 301 gctgggaccagagcttctgaagctgtgaagaaactgctgagcgctatcacacttgg 360
Db 314 GCTGGGACCAGAGCTTCTGAAGCTGTGAAGAACTGCTGGAGCGCTATCACACTTGGT 373
Qy 361 gtgacatcaagataaaaagagcgaggtggtgagtggaagatgatgctcctcctccc 420
Db 374 GTGACATCAAGATAAAGAGCGGAGGTGGATGGGATGGAAGATCATCTCTATCTCCC 433
Qy 421 tgctgaaacctgttctaccaaattatagatcaaatgccataaagttagtgaacctgtgaa 480
Db 434 TGCTGAAACCTGTCTTACCAATTATAGATCAAAATGCCCTAAATGTAGTGACCCGTGAA 493
Qy 481 aaggacaaaataaagcaatgaataca 505
Db 494 AAGGACAAATAAAGCAATGAATAA 518

RESULT 3
BG545709 529 bp mRNA EST 04-APR-2001
LOCUS 60257297F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4701042 5',
DEFINITION mRNA sequence.
ACCESSION BG545709
VERSION BG545709.1 GI:13544374
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 529)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1535 row: j column: 19
High quality sequence stop: 513.
FEATURES
source
Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4701042"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctggcc); Site_2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCCGACATG-dr(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 149 a 114 c 118 g 148 t
ORIGIN
Query Match 99.2%; Score 503; DB 155; Length 529;
Best Local Similarity 100.0%; Pred. No. 1.4e-140;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccaagtgaaccactgcttggtgatttgcctgattttctgatttttaaacctcctg 60
Db 13 GGCAGTGGAACCACTGGCTTGGTGATTTCCTAGATTTTCTGATTTTAAACTCTGT 72

```

QY 61 aaaaataccagataactgtcatgaagctggtaactatcttctctgtggtgaccatcag 120  
|||||  
Db 73 AAAAAATCCAGATAACTGTGATGAAGCTGGTAACATATCTTCTGCTGCTGACCATCAG 132  
|||||  
QY 121 cctttgtagtactgtctactgctcttcctcatcaaaaagtgcccccttctgttgacaa 180  
|||||  
Db 133 CCTTGTAGTTACTGCTACTGCTTCCCTCATCAACAAAAGTGCCCTTCTGTTGACAA 192  
|||||  
QY 181 gtgggacaccttacccttgacaaacttctcctttatggtatccattaaagctctctct 240  
|||||  
Db 193 GTTGGGACCTTACCCTGTGACACATCTTCCCTTATGATCCATTAAGCTTCTTCT 252  
|||||  
QY 241 gaaaactctgggcatctctgttgagacacctgtgaggggctgaaggaagtgtgtaataga 300  
|||||  
Db 253 GAAAACCTCTGGCATTTCTGTGTAGCACCTTGTGTGGAGGGCTAAGGAAGTGTGTAATGA 312  
|||||  
QY 301 gctgggacagagcttctgaagctgtgaagaaactctggaggcgctatcaacttgggt 360  
|||||  
Db 313 GCTGGGACCAAGAGCTTCTGAAGCTGTGAAGAACTGCTGGAGCGCTATCACACTTTGGT 372  
|||||  
QY 361 gtgacatacaagataaagagcgaggtgtgagtgaggatggaagatgagtctctatctccc 420  
|||||  
Db 373 GTGACATCAGATAAAGAGCGGAGGTGGATGGGATGGAAGATGATGCTCTATCTCTCC 432  
|||||  
QY 421 tgcctgaacctgtctaccataatagatcaaatgcccctaaatgtagtgtgacctgtgaa 480  
|||||  
Db 433 TGCTGAAACCTGTCTTACCAATATATAGATCAAAATGCCCTAAATGTPAGTGACCCGTGAA 492  
|||||  
QY 481 aaggacaaaataaagcaatgaata 503  
|||||  
Db 493 AAGGACAAATAAAGCAATGAATA 515  
|||||

## RESULT 4

BE550203/c 509 bp mRNA EST 10-AUG-2000  
LOCUS 7b50b11.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3231645 3',  
mRNA sequence.

ACCESSION BE550203

VERSION BE550203.1 GI:9791895

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 509)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

[infoimage.llnl.gov](mailto:infoimage.llnl.gov)

Seq primer: -40UP from Gibco

High quality sequence stop: 483.

Location/Qualifiers

## FEATURES

## source

1. .509

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3231645"

/clone\_lib="NCI\_CGAP\_Lu24"

/tissue\_type="carcinoid"

/lab\_host="DH10B"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized

library NCI\_CGAP\_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo.  
BASE COUNT 145 a 115 c 113 g 136 t  
ORIGIN

Query Match 97.5%; Score 494.4; DB 137; Length 509;  
Best Local Similarity 99.8%; Pred. No. 5.6e-138;  
Matches 495; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 ccaactgctggtagatttttctgtagatttttctgatttttaaaactcctgaaaaataatccc 71

Db 500 CCACTGGCTTGGGGATTTTGTAGATTTTCTGATTTTAAACTCTCTGAAAAATATATCCC 441

QY 72 agataaactgtcatgaagctgtgtaactatcttctgctggtgaccatcagcctttttagtt 131

Db 440 AGATAAAGCTCATGAAGCTGGTAACATATCTTCTGCTGGTGACCATCAGCCTTTGTAGTT 381

QY 132 actctgactgcttctctcatcaaaaagtgccttctctgttgacaagtggcaacctt 191

Db 380 ACTCTGTACTGCTTCTCTCATCAACAAAGTGCCCTTCTCTGTTGACAAGTTGCACCTT 321

QY 192 tacctctggacaacattcttcttcttctggtatccatcaaaagcttctctgaaaaactctgg 251

Db 320 TACCTCTGGACACATTTCTTCCCTTTATGGATCCATTAAGCTTCTCTGAAAACTCTGG 261

QY 252 gcattctgttgagcaccttctgcttctgaggggctgaagagtggtgtaaatgagctgggaccag 311

Db 260 GCATTTCTGTTGAGCACCTTGTGAGCGGGCTAAGGAAGTGTGTAATGAGCTGGGACCAG 201

QY 312 aggtcttgaaagtgtgaagaaactgctggaggcgctatcaacacttgggtgacatacaag 371

Db 200 AGGCTTCTGAAGCTGTGAAGAAACTGCTGGAGCGCTATCACACTTGGTGTGACATCAAG 141

QY 372 ataaagagcgagtgatggggatggaagatgatctctatctcctcctcctcctggaacc 431

Db 140 ATAAAGAGCGGAGGTGATGGGATGGAAGATGATGCTCTCTATCTCTCTGCTGAAACC 81

QY 432 tgttctaccattatagatcaaatgccctaaatgtagtgcctgtgaaaggacaaata 491

Db 80 TGTCTTACCAATTATAGATCAAAATGCCCTAAATGTAGTAGCCCGTGAAGGACAAATA 21

QY 492 aagcaatgaatacatt 507

Db 20 AAGCAATGAATACATT 5

## RESULT 5

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

AI355302 493 bp mRNA EST 04-JAN-1999

qt57e07.x1 Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone

IMAGE:1952100 3', mRNA sequence.

AI355302

AI355302.1 GI:4095455

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 493)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([infoimage.llnl.gov](mailto:infoimage.llnl.gov)) for further information.

Seq primer: -40UP from Gibco

LOCUS	DEFINITION	ACCSSION	VERSION	EST	IMAGE
6	60356877f3l NIH_MGC_77	825 bp.	03-APR-2001		
6G540454	mRNA sequence.	6G540454.1	GI:13532687		
6G540454					

```

Qy 420 ctgctgaacactgtctaccattatagatcaaatgcctctaaatgtagtgaccctgtga 479
      |||||||
Db 433 CTGCTGAACCTGTTCTACCAATATAGATCAATGCCCTAAATGTAGTGACCCGTGA 492

Qy 480 aaaggacaaataaagcaatgaatac 504
      |||||||
Db 493 AAGGACAAANTAAGCAATGAATAC 517

RESULT 7
BG541018 757 bp mRNA EST 03-APR-2001
LOCUS 602570177F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4694514 5',
DEFINITION mRNA sequence.
ACCESSION BG541018
VERSION BG541018.1 GI:13533251
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 757)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI518 row: j column: 19
High quality sequence stop: 525.
Location/Qualifiers
1. .757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4694514"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctcgcc); Site 2: SfiI (ggccattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT 215 a 172 c 193 g 177 t
ORIGIN

Query Match 95.18; Score 482; DB 155; Length 757;
Best Local Similarity 99.68; Pred. No. 3.4e-134;
Matches 504; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 ggcgaagtgaaccactggcttggtgattgtctagattttctgatttttaaaactctgt 60
      |||||||
Db 13 GGCAAGTGGACCACTGGCTGGTGGATTTGCTAGATTTTCTGATTTTAAACTCTGT 72

Qy 61 aaaaatccccagataaactgtcatgaagctggtaactatcttctgtctgtggaacatacg 120
      |||||||
Db 73 AAAAATATCCAGATAACTGTCATGAAGCTGTAACCTATCTTCTGCTGGTGACCATCAG 132

Qy 121 cctttgtagttaactgtctactgtctctctcatcaaaagcgcccttctctgttgacaa 180
      |||||||
Db 133 CCTTTGTAGTTACTCTGCTACTGCTTCTCCTCATCAACAAAGTGCCCTTCTCTGTGACAA 192

```

```

Qy 181 gttggacacctttacctctgtgacaacattcttccctttatgtatcattaaagctttct 240
      |||||||
Db 193 GTTGGACCTTTACTCTGGACACATTTCTCCCTTTATGGATCCATTAAAGCTTCTTCT 252

Qy 241 gaaaactctggcatttcttctgttgagcaccttgtgaggggctaaaggaagtgttaaatga 300
      |||||||
Db 253 GAAAACTCTGGGCATTTCTGTTGAGCACCTTGTGGAGGGGCTAAGGAAGTGTGTAATGA 312

Qy 301 gctggagcaggagcttctctgaagctgtgaagaaactctgtgaggcgctatcacacttggt 360
      |||||||
Db 313 GCTGGGACCAAGAGGCTTCTGAAGCTGTGAAGAACTGCTGGAGGCGCTATCACACTTGGT 372

Qy 361 gtgacatcaaatgataaaagcgaggtgtgattgggatggaagatgatgctctatctctccc 420
      |||||||
Db 373 GTGACATCAAGATAAAGAGCGGAGTGGATGGGATGGAAGATGATGCTCTATCTCTCCC 432

Qy 421 tgcctgaacacctgtctaccatattatagatcaaatgcctctaaatgtag-tgacccgtga 479
      |||||||
Db 433 TGCCTGAACCTGTTCTACCAATTTATAGATCAAAATGCCCTAAATGTAGTTGACCCGTGA 492

Qy 480 a-aaggacaaataaagcaatgaatac 504
      |||||||
Db 493 ACAAGGACAAATAAAGCAATGAATAC 518

RESULT 8
BG484055 825 bp mRNA EST 21-MAR-2001
LOCUS 602504659F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4617999 5',
DEFINITION mRNA sequence.
ACCESSION BG484055
VERSION BG484055.1 GI:13416334
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 825)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI375 row: f column: 16
High quality sequence stop: 491.
Location/Qualifiers
1. .825
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4617999"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctcgcc); Site 2: SfiI (ggccattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT 221 a 202 c 192 g 210 t
ORIGIN

```

<hr/>					
Query Match            94.5%;   Score 479.2;   DB 154;   Length 825;					
Best Local Similarity   98.0%;   Pred. No. 2.5e-133;					
Matches 496;   Conservative     0;   Mismatches 8;   Indels     2;   Gaps					
<hr/>					
Qy	1	ggcaagtggaaaccactggttggtgattttgcagatcttctgatcttttaaacctcctg	60		
Db	13	GGCAAGTGGAAACACACATGCGCTTGGTGGATTTCGTAGATTTTCTGATTTTAAACTCCTG	72		
Qy	61	aaaatatccacagataaactgcatgaagctggttaactatcttcctgctggtagaccatcag	120		
Db	73	AAAAATATCCAGATAAATGTCATGAAGCTGGTAAGTATCTTCCTGCCTGGTGACCATCAG	132		
Qy	121	cctttgtagttaacctgctgactgcttcctccotcalcaacaagtgccccttcctgttgacaa	180		
Db	133	CCTTTGTAGTTACTCTGCTGACTGCCCTTCCTCATCAACAAGTGCCCCCTCCTGTTGACAA	192		
Qy	181	gttggcacctttacacctctggacaacattcttcctcttatgatacattaaagctctctct	240		
Db	193	GTTGGCACCTTTACCTCTGSGAACAAATCTTCCCCTTTAAGGATCCATTAAGCTTCTTCT	252		
Qy	241	gaaacctctgggcattctctgttgagcaccttgtggaggcgctaagaagtggttaaatga	300		
Db	253	GAAACTCTGGGCATTTCTGTGAGCACTGTGTGAGGGGCTAAGGAAGTGTGTAAC TGA	312		
Qy	301	gctgggaccagaggtctctgaagctgtgaagaaactctggaggcgctacacacttggt	360		
Db	313	GCTGGACACAGAGGCTCTGAACTGTGAGAAACTGCTGGAGGCGCTATCACACTTGGT	372		
Qy	-361	gtgacatacaagataaagagcgaggtggatgggatg--aagatgatgctctatcctc	418		
Db	-373	GTGACATCAAGATCAAGAGCGAGCGTGGTGGGATGGCACAGATGATGCTCTATCCTC	432		
Qy	419	ctgctgtaaaccgttctaccattatagatacaaatgccataaaatgtagtgaccctgtg	478		
Db	-433	CCTGGCTGAAACCTGTTCTACCAATTAAGATCAAAATGCGCTAAAATGAGTAGCACCGTG	492		
Qy	479	aaaggagcaataaagaacatgaatc	504		
Db	493	ACAAGGCAAAATAAAGCAATGAATAC	518		
<hr/>					
RESULT	9				
Ail193988/c					
LOCUS					
DEFINITION	Ail193988	474 bp	mRNA	EST	13-OCT-1998
	ge81c04.x1 Soares_fetal_lung_NDHL19W Homo sapiens cDNA clone				
	IMAGE:I1745382 3', mRNA sequence.				
ACCESSION	Ail193988				
VERSION	Ail193988.1	GI:3745197			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
JOURNAL	Tumor Gene Index				
COMMENT	Unpublished (1997)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-remail.nih.gov				
	This clone is available royalty-free through LLNL ; contact the				
	IMAGE Consortium (infoimage.llnl.gov) for further information.				
	Seq primer: -400P from Gibco.				
FEATURES	Location/Qualifiers				
source	1..474				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:I1745382"				
	/clone_lib="Soares_fetal_lung_NDHL19W"				
	/dev_stage="19 weeks"				
	/lab_host="DH10B (ampicillin resistant)"				
	/note="Organ: lung; Vector: pTZ19 (Pharmacia) with a				
	modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st				



cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov

Seq primer: -40UP from Gbco.

## FEATURES

source  
1. .461  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3566243"  
/clone\_lib="NCI-CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI-CGAP\_Lu5 was prepared, and ss circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 126 a 102 c 103 g 129 t 1 others  
ORIGIN

Query Match 90.3%; Score 458; DB 144; Length 461;  
Best Local Similarity 99.8%; Pred. No. 4.9e-127;  
Matches 458; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 47 tttttaactctgaaaaatcccagataactgtcatgaagctggaactatcttctcg 106  
|||||  
Db 461 TTTTAACTCTGANAATATCCAGATAACTGTCATGAAGCTGGTAACATCTTCTCG 402  
Qy - 107 ctggtgaccatcagccttctgtagtactctgctactgctctcctcctcaacaaagtgc 166  
|||||  
Db 401 CTGCTGACCATCAGCCTTTGTAGTTACTCTGCTACTGCTTCTCTCATCAACAAAGTGCCC 342  
Qy 167 ctctcttgacaagtggaccctttacctctggacacattctccctttatgatcca 226  
|||||  
Db 341 CTTCCTGTGTGACAGTGGACACCTTTACTCTGGACAACTCTTCCCTTTATGGATCCA 282  
Qy 227 ttaagctctcttgaaaactctggcattctctgtgagcaccttctgaggggctaaagg 286  
|||||  
Db 281 TTAAGCTTCTTGAAACTCTGGCATTTCTGTGAGCACCTTGTGAGGGGCTAAGG 222  
Qy 287 aagtgtataatgagctgggaccagagcttctgaagctgtgaagaaactgctggagcgc 346  
|||||  
Db 221 AAGTGTGTAATGAGCTGGGACCAAGAGCTTCTGAAGCTGTGAAGAACTGCTGGAGCG 162  
Qy 347 ctatcacacttggtgacatcaagataaagcggaggtgagtgaggatggaagatgat 406  
|||||  
Db 161 CTATCACACTTGGGTGACATCAAGATAAAGCGGAGGTGGATGGAGATGAT 102  
Qy 407 gctcctatcctccctgctgaaactgttctaccaattatagatcaaatgcctcaaatg 466  
|||||  
Db 101 GCTCCTATCCTCCCTGCTGAACTGTTCTACCAATTTATAGATCAATGCCCTAAATG 42  
Qy 467 tagtgaccctgtgaaaaggacaaataaagcaatgaatata 505  
|||||  
Db 41 TAGTGACCCGTGAAAGGACAAATAAAGCAATGAATACA 3

RESULT 11  
BE348591/c  
LOCUS  
DEFINITION hc72d05.x1 NCI-CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3152265 3',  
mRNA sequence.  
ACCESSION BE348591  
VERSION BE348591.1 GI:9260444  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 462)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov

## FEATURES

## source

1. .462  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3152265"  
/clone\_lib="NCI-CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI-CGAP\_Lu5 was prepared, and ss circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 127 a 102 c 103 g 130 t  
ORIGIN

Query Match 90.1%; Score 456.8; DB 166; Length 462;  
Best Local Similarity 99.6%; Pred. No. 1.1e-126;  
Matches 458; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 46 atttttaactctgaaaaatcccagataactgtcatgaagctggttaactatcttct 105  
|||||  
Db 462 ATTTTAACTCTCGAAAAATATCCACATAACTGTCATGAAGCTGGTAACATCTTCTCT 403  
Qy 106 gctggtgaccatcagccttctgtagtactctgctactgctctcctcctcaacaaagtgc 165  
|||||  
Db 402 GCTGCTGACCATCAGCCTTTGTAGTTACTTGTCTACTGCTTCTCCATCAACAAAGTGCC 343  
Qy 166 ccttctcttgacaagttggcacctttacacctctggacaaacattctccctttatgatcc 225  
|||||  
Db 342 CCTTCTCTTGACAAAGTTGGCACCTTTTACCTCTGGACAAACATTTCTCCTTTATGATGCC 283  
Qy 286 attaaagctctcttgaaaaactctggccattctctgttagcacctcttgagggggctaag 285  
|||||  
Db 282 ATTAAGACTTCTTCTGAAAACTCTGGGCATTTCTGTTGAGCACCTTGTGGAGGGCTAAG 223  
Qy 286 gaagtgtataatgagctgggaccagagccttctgaaactgtgaagaactgctggagggc 345  
|||||  
Db 222 TAAGTGTGTAATGAGCTGGGACCAAGAGCCTTCTCAAGCTGTGAAGAAACTGCTGGAGGC 163  
Qy 346 gctatcacacttggtgtgacatcaagataaagcggagggtggatggggatggaagatga 405  
|||||  
Db 162 GCTATCACACTTGGTGTGACATCAAGATAAAGCGGAGGTGGATGGGATGGAAGATGA 103  
Qy 406 tgcctctatcctcctgctgctgaaaccttctaccatattatagatcaaatgctccctaaat 465  
|||||  
Db 102 TGCTCTATCTCTCCCTGCCTGAAACCTGTTCTTACCAATTTATAGATCAATGCCCTAAAA 43  
Qy 466 gtagtgaaccctgtgaaaaggacaaataaagcaatgaatata 505

```

Db 42 GTAGTGACCCGTGAAAGGCAATAAAGCAATGGATACA 3
|||||
RESULT 12
LOCUS BG547781 817 bp mRNA EST 04-APR-2001
DEFINITION 602575659F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4703631 5',
mRNA sequence.
ACCESSION BG547781
VERSION BG547781.1 GI:13546446
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM1542 row: f column: 16
High quality sequence stop: 470.
FEATURES
source
1. 817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4703631"
/lab_host="NIH-MGC-77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggccctcgcc); Site:2: SfiI (ggccatagccc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGCGGCGGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."
BASE COUNT 209 a 244 c 196 g 168 t
ORIGIN
Query Match 89.2%; Score 452; DB 155; Length 817;
Best Local Similarity 99.8%; Pred. No. 3.8e-125;
Matches 463; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 gcaagtggaaacacactgctggtgatttctgctgaattttctgatttttaaacctcgt 60
Db 12 GCGAAGTGGAAACACACGCTTGTTGGGA-TTTTCTAGATTTTCTGATTTTAAACTCTG 70
QY 61 aaaaatccccagataactgtcatgaagctgttaactatcttcctgctggtgacctcag 120
Db 71 AAAAATATCCAGATAACGTCTATGAAGCTGTAACATATCTTCCTGCTGGTGAACATCAG 130
QY 121 cctttgtactctgctactgcttctctctctcaacaagaagtgcccttctctgttgaca 180
Db 131 CCTTTCTACTTCTGCTACTGCTTCTCTCTCTCATCAACAAGTGCCCTCTCTGTGACA 190
QY 181 gttggacatttacctctgttgacaacattctctctttatgattccattaaagctttct 240
Db 191 GTTGGACCTTTTACCTCTGGACAACATTTCTCCCTTTTATGATCCATTAAGCTTCTCT 250
QY 241 gaaaactctgggcatttctgttgagcaactgtggaaggggtaagggaagtgtgataatga 300
|||||
us-09-700-770-1.lst
Db 251 GAAAACTCTGGGCATTTCTGTTGAGCACCTTGTGGAGGGCTAAGGAAGTGTCTAAATGA 310
QY 301 gctggaccagagcttctgaagctgtgaagaaactctgagggcgtatcacacttggt 360
Db 311 GCTGGGACCAAGAGCTTCTTGAAGCTGTGAAGAACTGCTGGAGCGCTATCACACTTGGT 370
QY 361 gtgacatcaagataaaagcgaggtggtggtggtggtggtggtggtggtggtggtggtggt 420
Db 371 GTGACATCAAGATAAAGAGCGGAGTGGATGGAAGATGATGCTCTCTCTCTCTCTCC 430
QY 421 tgctgaaactgtcttaccacattatagatcaaatgcctctaaa 464
Db 431 TGCTGAAACCTGTTCTTACCAATATAGATCAAAATGCCCTAAAA 474
|||||
RESULT 13
LOCUS N92483 510 bp mRNA EST 05-APR-1996
DEFINITION zbl2g04.s1 Soares_fetal_lung_NBHL19W Homo sapiens cDNA clone
IMAGE:301878 3' similar to contains element PTR5 repetitive element
; mRNA sequence.
ACCESSION N92483
VERSION N92483.1 GI:1264792
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ETPPrimer
High quality sequence stop: 354.
FEATURES
source
1. 510
/organism="Homo sapiens"
/db_xref="GDB:1246809"
/db_xref="taxon:9606"
/clone="IMAGE:301878"
/lab_host="Soares_fetal_lung_NBHL19W"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTCTTTTCTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Ronaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W."
BASE COUNT 141 a 120 c 112 g 137 t
Query Match 88.7%; Score 449.8; DB 187; Length 510;
Best Local Similarity 96.2%; Pred. No. 1.5e-124;

```

was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5'-TGTTCAACTGTAAGTGGAGCGCGCTTAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT	126 a	102 c	103 g	133 t
ORIGIN				
Query Match	88.08;	Score 446;	DB 14;	Length 464;
Best Local Similarity	99.98;	Pred. No. 2e-123;		
Matches 457;	Conservative	0;	Mismatches	0;
Indels	1;	Gaps	1;	
Qy	48	ttttaaactctgaaaaatattccagataactgcataagctgtaactatctctgc	107	
Db	461	TTTTTAAACTCTGAAAATATCCCGATACTGTCATGAAGTGTAACTATCTTCTGC	402	
Qy	108	tggtagacatcagcctttagttactctgctactgctctctctcatcaaaaaagtcccc	167	
Db	401	T-GTGACCATCAGCCTTTGTAGTTACTTGCTACTGCTCTCTCATCAACAAAGTGCCCC	343	
Qy	168	tctctgtgacaagtctggcacctttactctctggacaacattctccctttatgatccat	227	
Db	342	TTCTCTGTTGACAGTTGGCACCTTTAGCTCTCGGACAACTTCTTCCCTTTATGATGCAT	283	
Qy	228	taaagctctctgaaacactcggcattctctgttgagacactgtgagggcgtaagga	287	
Db	282	TAAAGCTTCTTCTGAAAACCTCGGGCATTTCTGTGAGCACCTTGTGAGGGGCTAAGGA	223	
Qy	288	agtggtataatgagctgggacagagctctctgaagctgtgaagaaactgctggagcgcg	347	
Db	222	AGTGTGTAATGACCTGGGACCAGAGGCTTCTGAAGCTGTGAAGAACTGCTGGAGGGCG	163	
Qy	348	tatcacacttggtgtgacataagaataagagcggaggtggatgggaatggaatgatg	407	
Db	162	TATCACACTTGGTGTGACATCAAGATAAAGACGCGAGGTGGATGGGATGGAAGATGATG	103	
Qy	408	ctctatctctctgcctgaaacctgttctaccaattatagatcaaatgcccataaatgt	467	
Db	102	CTCTATCTCTCCCTGCCGTGAACCTGTCTACCAATATATAGATCAAAATGCCCTAAAATGT	43	
Qy	468	agtgaccctgtaaaggacaaaataaagcaatgaataca	505	
Db	42	AGTGACCCGTGAAAGGACCAATAAAGCAATGAATACA	5	
RESULT 15				
AI193016/c				
LOCUS	AI193016	445 bp	mRNA	EST
DEFINITION	q69a09.x1 Soares_fetal_lung_NbHL19W	Homo sapiens	CDNA clone	29-OCT-1998
	IMAGE:1744216	3', mRNA sequence.		
ACCESSION	AI193016			
VERSION	AI193016.1	GI:3744225		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 445)			
TITLE	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.			
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
JOURNAL	Tumor Gene Index			
COMMENT	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: ccapbs-r@mail.nih.gov			
	This clone is available royalty-free through LNL; contact the			
	IMAGE Consortium (info@image.llnl.gov) for further information.			
	Insert Length: 303 Std Error: 0.00			
	Seq primer: -40UP from Gibco.			
FEATURES	Location/Qualifiers			

```
source 1. .445
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:174216"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TCTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Patima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."

BASE COUNT 120 a 102 c 99 g 124 t
ORIGIN

Query Match 87.8%; Score 445; DB 17; Length 445;
Best Local Similarity 100.0%; Pred. No. 4e-123;
Matches 445; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 tgaataatccccgaataactgtcatgaagctggtaactatctctcgtggtgaccatc 118
Db 445 TGAATAATCCAGATAACTGTATGAAGCTGGTAACTATCTCTCGTGGTGACCATC 386

Qy 119 agcctttagttagttagttagttagttagttagttagttagttagttagttagt 178
Db 385 AGCCTTTGTAGTACTGTACTGTACTGTACTGTACTGTACTGTACTGTACTGT 326

Qy 179 aagttggcacctttacacctgtggacaacattctccctttatggatccattaaagcttct 238
Db 325 AAGTTGGCACCTTTACCTCTGGACAACATCTTCCCTTTATGGATCCATTAAAGCTTCTT 266

Qy 239 ctgaataactctggcattctgttgacacctgtgtgaggggctaagaagtggtgtaaat 298
Db 265 CTGAATAACTCTGGCATTCTGTGTGACACCTTGTGGAGGGGCTAAGGAAGTGTGTAAT 206

Qy 299 gagctgggaccagaggcttctgaagctgtgaagaactgctggaggcgctatcacacttg 358
Db 205 GAGCTGGGACCAGAGGCTTCTGAAGCTGTGAAGAACTGCTGGAGGCGCTATCACACTTG 146

Qy 359 gtgtgacataaagataaagcggaggtggatggggaatgaagatgatgctctctctc 418
Db 145 GTGTGACATCAAGATAAAGAGCGGAGGTGGATGGGATGGGAAGATGATGCTCTATCCTC 86

Qy 419 cctgcctgaacctgttctaccattatagatcaaatgcctaaatgtagtgacccgtg 478
Db 85 CCTGCCTGAAACCTGTCTACCAATTATAGATCAATGCCCTAAATGTAGTGACCCGTG 26

Qy 479 aaagagacaataaagcaatgaata 503
Db 25 AAGAGGACAATAAAGCAATGAATA 1
```

Search completed: November 17, 2001, 13:33:57  
Job time: 3396 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: November 17, 2001, 12:51:12 ; Search time 222.76 Seconds  
(without alignments)  
1429.099 Million cell updates/sec

Title: US-09-700-770-1  
Perfect score: 507  
Sequence: 1 ggcaagtgaaccactggct.....aataaagcaatgaatacatt 507

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	507	100.0	507	19 AAV38069	Human lung tissue
2	507	100.0	507	20 AAX99424	LU103 specific pol
3	507	100.0	507	21 AA229721	Human lung specific
4	506	99.8	531	21 AA298174	Human signal pepti
5	499.4	98.5	519	19 AAV38070	Human lung tissue
6	499.4	98.5	519	20 AAX99425	LU103 specific pol
7	325.4	64.2	357	20 AAV86241	EST clone AA246
8	263	51.9	263	19 AAV38067	Human lung tissue
9	263	51.9	263	20 AAX99422	LU103 specific pol
10	263	51.9	269	19 AAV38066	Human lung tissue
11	263	51.9	269	20 AAX99421	LU103 specific pol

12	225	44.4	225	19 AAV38068	Human lung tissue
13	225	44.4	225	20 AAX99423	LU103 specific pol
14	58	11.4	60	16 AAX22859	Human gene signatu
15	56.4	11.1	225	19 AAV54618	LU105 specific pol
16	56.4	11.1	519	19 AAV54621	LU105 polypeptide
17	56.4	11.1	543	21 AA298173	Human signal pepti
18	56.4	11.1	543	21 AA229723	Human lung specific
19	56.4	11.1	562	19 AAV54620	LU105 specific con
20	56.4	11.1	570	21 AAV65103	Membrane-bound pro
21	56.4	11.1	570	22 AAF44249	Human PRO1245 (UNQ
22	45	8.9	936	22 AAF58252	Oligonucleotide D1
23	45	8.9	936	22 AAF58254	Oligonucleotide D1
24	45	8.9	936	22 AAF58257	Oligonucleotide D1
25	45	8.9	936	22 AAF58259	Oligonucleotide D2
26	45	8.9	936	22 AAF58262	Oligonucleotide D2
27	45	8.9	938	22 AAF58255	Oligonucleotide D1
28	42.2	8.3	936	22 AAF58252	Oligonucleotide D1
29	42.2	8.3	936	22 AAF58254	Oligonucleotide D1
30	42.2	8.3	936	22 AAF58257	Oligonucleotide D1
31	42.2	8.3	936	22 AAF58259	Oligonucleotide D1
32	42.2	8.3	936	22 AAF58262	Oligonucleotide D2
33	42.2	8.3	938	22 AAF58255	Oligonucleotide D1
34	34	6.7	1324	21 AAC45583	Arabidopsis thalia
35	34	6.7	1325	21 AAC33679	Arabidopsis thalia
36	34	6.7	1431	21 AAC46348	Arabidopsis thalia
37	34	6.7	1435	21 AAC40039	Arabidopsis thalia
38	33.4	6.6	1943	21 AAC47114	Arabidopsis thalia
39	33.4	6.6	5019	21 AAC45709	Arabidopsis thalia
40	32.6	6.4	3471	18 AAY71761	PSKH-1 cDNA. Homo
41	32.4	6.4	2815	20 AAX06817	Chlamydia pneumoni
42	32.4	6.4	2950	21 AAX30849	Chlamydia antigen
43	32.4	6.4	3000	20 AAX06828	Chlamydia pneumoni
44	32.4	6.4	8689	21 AAA97996	Human T gene cDNA.
45	32.4	6.4	8949	21 AAA07840	Human homologue of

ALIGNMENTS

RESULT 1  
AAV38069  
ID AAV38069 standard; cDNA; 507 BP.  
XX AC AAV38069;  
XX 14-SEP-1998 (first entry)  
XX Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:4.  
XX Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FT CDS 83..364  
FT /\*tag= a  
FT /\*product= "LU103"  
XX WO9820143-A1.  
XX 14-MAY-1998.  
XX 05-NOV-1997; 97WO-US20680.  
XX 05-NOV-1996; 96US-0744211.  
XX (ABBO ) ABBOTT LAB.  
XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
XX Kratochvill JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
XX WPI; 1998-286957/25.  
XX P-PSDB; AAW62068.

XX Lung tissue derived polynucleotide LU103 - useful to detect,  
 PT diagnose, stage, monitor, prognosis, prevent, treat or determine  
 XX pre-disposition to lung disease, e.g. lung cancer  
 XX  
 PS Claim 1; Page 67; 86pp; English.  
 XX  
 CC The present sequence represents a polynucleotide specific for lung  
 CC tissue gene LU103. A method has been developed for detecting the  
 CC presence of a target LU103 polynucleotide in a test sample, comprising:  
 CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,  
 CC and (b) detecting the target LU103 polynucleotide in the test sample,  
 CC where the LU103 polynucleotide has at least 50% identity to the 269,  
 CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to  
 CC AAV38070. The methods and products of the present invention may be used  
 CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or  
 CC determine the predisposition diseases and conditions of the lung, e.g.  
 CC lung cancer.  
 XX  
 XX Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;

Query Match 100.0%; Score 507; DB 19; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-156;  
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcaagtgaaccactggctgtggatttgcctgtagattttctgatttttaaaactctg 60  
 Db 1 ggcaagtgaaccactggctgtggatttgcctgtagattttctgatttttaaaactctg 60  
 QY 61 aaaaatccccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120  
 Db 61 aaaaatccccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120  
 QY 121 cctttgtagttactgctgactgcttctccatcaacaaagtgccttctctgttgacaa 180  
 Db 121 cctttgtagttactgctgactgcttctccatcaacaaagtgccttctctgttgacaa 180  
 QY 181 gttggacccttaccctggacacatttccctttatgagatgacattaaagtcttct 240  
 Db 181 gttggacccttaccctggacacatttccctttatgagatgacattaaagtcttct 240  
 QY 241 gaaactctggcattctgtgagcaccttgtggagggctgaaggaagtgttaaatga 300  
 Db 241 gaaactctggcattctgtgagcaccttgtggagggctgaaggaagtgttaaatga 300  
 QY 301 gctgggaccagagcttctgaagctgtgaagaaactgctggagggctatcacacttgg 360  
 Db 301 gctgggaccagagcttctgaagctgtgaagaaactgctggagggctatcacacttgg 360  
 QY 361 gtgacatcaagataaagagcggagggtggatgggagtggaagatgactcctatctccc 420  
 Db 361 gtgacatcaagataaagagcggagggtggatgggagtggaagatgactcctatctccc 420  
 QY 421 tgctgaaaccttctaccataattatagatcaaatgcccataaatgtagtgaccocgtgaa 480  
 Db 421 tgctgaaaccttctaccataattatagatcaaatgcccataaatgtagtgaccocgtgaa 480  
 QY 481 aaggacaaataaagcaatgaatacatt 507  
 Db 481 aaggacaaataaagcaatgaatacatt 507

RESULT 2  
 AAX99424  
 ID AAX99424 standard; DNA; 507 BP.  
 XX AC  
 XX AAX99424;  
 XX DT 19-OCT-1999 (first entry)  
 XX DE LU103 specific polynucleotide consensus sequence.

KW LU103; tumour; lung cancer; detection;  
 KW Fluorescent in situ hybridisation; ss.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 85..364  
 FT /\*tag= a  
 FT /product= "Sequence AAY28334 amino acid residue"  
 PN US939265-A.  
 PD 17-AUG-1999.  
 XX  
 XX 05-NOV-1997; 97US-0964725.  
 XX  
 XX 05-NOV-1997; 97US-0964725.  
 PR 05-NOV-1996; 96US-0744211.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
 XX  
 DR WPI: 1999-468402/39.  
 XX  
 XX P-PSDB; AAY28334.  
 XX  
 PT Polynucleotides useful for detecting, diagnosing and monitoring  
 PT diseases of the lung such as lung cancer  
 PS Claim 1; Column 47-49; 36pp; English.  
 XX  
 CC The 507 base pair sequence of a LU103 specific polynucleotide was  
 CC derived from a consensus of the isolated sequences from clones  
 CC 1235095, 1235531 and 1379417.  
 CC The polynucleotides and methods disclosed in the invention can be  
 CC useful for detecting, diagnosing, staging, monitoring or predicting  
 CC diseases and conditions of the lung, such as lung cancer.  
 CC The polynucleotides may be used to produce probes for use in  
 CC fluorescent in situ hybridization (FISH) technology to perform  
 CC chromosomal analysis and identify cancer specific alterations  
 CC such as deletions.  
 XX  
 XX Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;

Query Match 100.0%; Score 507; DB 20; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-156;  
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcaagtgaaccactggctgtggatttgcctgtagattttctgatttttaaaactctg 60  
 Db 1 ggcaagtgaaccactggctgtggatttgcctgtagattttctgatttttaaaactctg 60  
 QY 61 aaaaatccccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120  
 Db 61 aaaaatccccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120  
 QY 121 cctttgtagttactgctgactgcttctccatcaacaaagtgccttctctgttgacaa 180  
 Db 121 cctttgtagttactgctgactgcttctccatcaacaaagtgccttctctgttgacaa 180  
 QY 181 gttggacccttaccctggacacatttccctttatgagatgacattaaagtcttct 240  
 Db 181 gttggacccttaccctggacacatttccctttatgagatgacattaaagtcttct 240  
 QY 241 gaaactctggcattctgtgagcaccttgtggagggctgaaggaagtgttaaatga 300  
 Db 241 gaaactctggcattctgtgagcaccttgtggagggctgaaggaagtgttaaatga 300  
 QY 301 gctgggaccagagcttctgaagctgtgaagaaactgctggagggctatcacacttgg 360  
 Db 301 gctgggaccagagcttctgaagctgtgaagaaactgctggagggctatcacacttgg 360

Db 301 gctgggaccagagcttctgaagctggaagaaactgctggaggcgtatcacacttggt 360  
 Qy 361 gtaacatcaagataaagcgaggtgagtgaggatggaagatgagctctctctccc 420  
 Db 361 gtaacatcaagataaagcgaggtgagtgaggatggaagatgagctctctccc 420  
 Qy 421 tgctgaaactgcttaccacattatagatcaaatgcccataaaatgtagtaccctgaa 480  
 Db 421 tgctgaaactgcttaccacattatagatcaaatgcccataaaatgtagtaccctgaa 480  
 Qy 481 aaggacaaataaagcaatgaatacatt 507  
 Db 481 aaggacaaataaagcaatgaatacatt 507

## RESULT 3

AAZ29721  
 ID AAZ29721 standard; DNA; 507 BP.

XX AC AAZ29721;

XX DT 27-MAR-2000 (first entry)

XX DE Human lung specific gene Lngl01.

XX KW Lung Specific Gene; LSG; Lngl01; human; diagnostic marker;  
 KW prognosticate; lung cancer; diagnosis; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 FT 83..364  
 FT /\*tag= a  
 FT /product= "LSG Lngl01 protein"

XX X WO9960160-A1.

XX PD 25-NOV-1999.

XX PF 12-MAY-1999; 99WO-US10344.

XX PR 21-MAY-1998; 98US-0086212.

XX PA (DIAD-) DIADEXUS LLC.

XX PI Yang R, Macina RA, Sun Y;

XX X WPI; 2000-116320/10.

XX DR P-PSDB; AAY44456.

XX PT A new method for diagnosing, monitoring and staging lung cancer -

XX PS Claim 6; Page 33; 40pp; English.

XX CC The present sequence is a lung specific gene (LSG) (Lngl01) from human  
 CC clone ID 126758. The LSG has high level of tissue specificity for lungs  
 CC and is overexpressed in cancerous tissues. The sequence serves as a  
 CC diagnostic marker for detecting, monitoring, staging and prognosticating  
 CC lung cancer. The diagnosis involves comparing levels of LSG in samples  
 CC obtained from patient and normal control.

XX SQ Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;

## Query Match

Best Local Similarity 100.0%; Score 507; DB 21; Length 507;  
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggaagtggaaccactggctgggtgagtgagatgttttctgattttttaaactctg 60

Db 1 ggaagtggaaccactggctgggtgagtgagatgttttctgattttttaaactctg 60

Qy 61 aaaaataccccagataaactgcatgaagctggttaactatctctcgtggtgaccatcag 120

Db 61 aaaaataccccagataaactgcatgaagctggttaactatctctcgtggtgaccatcag 120  
 Qy 121 ccttctgtagttactctgctactgcttctctccatcaacaaagtgcctcttctgttgacaa 180  
 Db 121 ccttctgtagttactctgctactgcttctctccatcaacaaagtgcctcttctgttgacaa 180  
 Qy 181 gttggcacccttacccttggaacacattcttccctttatggtcattaaagtcttctt 240  
 Db 181 gttggcacccttacccttggaacacattcttccctttatggtcattaaagtcttctt 240  
 Qy 241 gaaactctgggcatttctgttgagcaacctgttgagggtgctaaagtggtgtaaatga 300  
 Db 241 gaaactctgggcatttctgttgagcaacctgttgagggtgctaaagtggtgtaaatga 300  
 Qy 301 gctgggaccagagcttctgaagctggaagaaactgctggaggcgtatcacacttggt 360  
 Db 301 gctgggaccagagcttctgaagctggaagaaactgctggaggcgtatcacacttggt 360  
 Qy 361 gtgacatcaagataaagcgaggtgagtgaggatggaagatgagctctctctccc 420  
 Db 361 gtgacatcaagataaagcgaggtgagtgaggatggaagatgagctctctctccc 420  
 Qy 421 tgctgaaacctgttaccacattatagatcaaatgcccataaaatgtagtaccctgtaa 480  
 Db 421 tgctgaaacctgttaccacattatagatcaaatgcccataaaatgtagtaccctgtaa 480  
 Qy 481 aaggacaaataaagcaatgaatacatt 507  
 Db 481 aaggacaaataaagcaatgaatacatt 507

## RESULT 4

AAZ98174  
 ID AAZ98174 standard; cDNA; 531 BP.

XX AC AAZ98174;

XX DT 11-MAY-2000 (first entry)

XX DE Human signal peptide containing protein HSP66 cDNA SEQ ID NO:200.

XX KW Human; signal peptide-containing protein; HSP66; diagnosis; cancer;  
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
 KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;  
 KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;  
 KW reproductive disorder; developmental disorder; arteriosclerosis;  
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;  
 KW muscular dystrophy; ss.

XX OS Homo sapiens.

XX PN WO2000000610-A2.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-US14484.

XX PR 26-JUN-1998; 98US-0090762.

XX PR 31-JUL-1998; 98US-0094983.

XX PR 01-OCT-1998; 98US-0102686.

XX PR 11-DEC-1998; 98US-0112129.

XX X (INCY-) INCYTE PHARM INC.

XX PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;

XX PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;

XX X WPI; 2000-160673/14.

XX DR P-PSDB; AAY87289.

XX New human signal peptide-containing proteins useful in treatment,  
 PT prevention and diagnosis of e.g. cancer, inflammation and  
 PT cardiovascular disease  
 XX  
 PS Claim 9; Page 289; 327pp; English.  
 XX  
 CC AA298109 to AA298242 encode AY87224 to AY87357 which represent the  
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have  
 CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,  
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can  
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders  
 CC associated with decreased activity or function of HSP. Antagonists of  
 CC HSP are used to treat or prevent disorders associated with increased  
 CC activity or function of HSP. Such diseases include cell proliferation  
 CC (including cancer), inflammation, cardiovascular, neurological,  
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,  
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anemia,  
 CC asthma, Crohn's disease, microbial or other infections, congestive or  
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP  
 CC nucleic acids can be used for the recombinant production of HSP, for  
 CC detecting HSP in standard hybridisation and amplification assays (for  
 CC triplex-forming or ribozyme therapeutics), in gene therapy, as antisense,  
 CC or genetic variations, and for chromosomal mapping. HSP are also used to  
 CC raise specific antibodies (Ab) and to screen for agonists and  
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic  
 CC antagonists, in competitive drug screens, and for purification of HSP  
 CC from natural sources.  
 XX  
 SQ Sequence 531 BP; 149 A; 115 C; 119 G; 148 T; 0 other;

Query Match 99.8%; Score 506; DB 21; Length 531;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-155;  
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 gcaagtgaaccactgcttgggtgattgctagattttctgatttttaaaactctga 61  
 DB 1 gcaagtgaaccactgcttgggtgattgctagattttctgatttttaaaactctga 60  
 QY 62 aaaatataccagataactgctgaagctggttaactattcttctgctggtgaccatcag 121  
 DB 61 aaaatataccagataactgctgaagctggttaactattcttctgctggtgaccatcag 120  
 QY 122 cttttagttactctgcttctctctctcaacaagaagtgcccttctctgttgacaag 181  
 DB 121 cttttagttactctgcttctctctctcaacaagaagtgcccttctctgttgacaag 180  
 QY 182 ttggacatttactcttggaacatttcttcttattgattccattaaagcttctctg 241  
 DB 181 ttggacatttactcttggaacatttcttcttattgattccattaaagcttctctg 240  
 QY 242 aaactctggcattctgttgaacacctgttgagggggctgaagtggttaaatgag 301  
 DB 241 aaactctggcattctgttgaacacctgttgagggggctgaagtggttaaatgag 300  
 QY 302 ctgggaccagagcttctgaagctgtgaagaactgctgagcgctatcacacttggtg 361  
 DB 301 ctgggaccagagcttctgaagctgtgaagaactgctgagcgctatcacacttggtg 360  
 QY 362 tgacatcaagataaagagcgagggtgagtgaggatggaagatgctctatccct 421  
 DB 361 tgacatcaagataaagagcgagggtgagtgaggatggaagatgctctatccct 420  
 QY 422 gctgaacctgttctaccaattatagatcaaatgccctaaatagtagcccggtgaaa 481  
 DB 421 gctgaacctgttctaccaattatagatcaaatgccctaaatagtagcccggtgaaa 480  
 QY 482 aggaacaataaagcaatgaatcatt 507  
 DB 482 aggaacaataaagcaatgaatcatt

Db 481 aggaacaataaagcaatgaatcatt 506  
 RESULT 5  
 ID AAV38070  
 XX AAV38070 standard; cDNA; 519 BP.  
 AC AAV38070;  
 XX 14-SEP-1998 (first entry)  
 DE Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:5.  
 XX Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.  
 XX Homo sapiens.  
 FH Key Location/Qualifiers  
 CDS 96..377  
 FT /\*tag= a  
 FT /product= "LU103"  
 XX WO9820143-A1.  
 PD 14-MAY-1998.  
 XX 05-NOV-1997; 97WO-US20680.  
 PR 05-NOV-1996; 96US-0744211.  
 XX (ABBO ) ABBOTT LAB.  
 XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
 PI Kratochvill JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
 XX WPI: 1998-286957/25.  
 DR P-PSDB; AAW62068.  
 XX Lung tissue derived polynucleotide LU103 - useful to detect,  
 PT diagnose, stage, monitor, prognosis, prevent, treat or determine  
 PT pre-disposition to lung disease, e.g. lung cancer  
 XX Claim 1; Page 68; 86pp; English.  
 CC The present sequence represents a polynucleotide specific for lung  
 CC tissue gene LU103. A method has been developed for detecting the  
 CC presence of a target LU103 polynucleotide in a test sample, comprising:  
 CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,  
 CC and (b) detecting the target LU103 polynucleotide in the test sample,  
 CC where the LU103 polynucleotide has at least 50% identity to the 269,  
 CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to  
 CC AAV38070. The methods and products of the present invention may be used  
 CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or  
 CC determine the predisposition diseases and conditions of the lung, e.g.  
 CC lung cancer.  
 XX Sequence 519 BP; 140 A; 113 C; 117 G; 149 T; 0 other;  
 SQ  
 Query Match 98.5%; Score 499.4; DB 19; Length 519;  
 Best Local Similarity 99.8%; Pred. No. 2.1e-153;  
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 gtggaacactgcttgggtgatttctgctagattttctgatttttaaaactcctgaaaaa 65  
 DB 19 gtggaacactgcttgggtgatttctgctagattttctgatttttaaaactcctgaaaaa 78  
 QY 66 tatccagataaactgtcatgaagctgaagctgtaactattcttctgctgggacatcagcttt 125  
 DB 79 tatccagataaactgtcatgaagctgtaactattcttctgctgggacatcagcttt 138  
 QY 126 gtagtactctgctactgcttctctcatcaacaagaagtgcccttctctgttgacaagtgg 185  
 DB 126 gtagtactctgctactgcttctctcatcaacaagaagtgcccttctctgttgacaagtgg



Db 139 gtagttactctgtactgcttccctcatcaacaagtgcccttctctgttgacaagtgg 198  
QY 186 cacccttacccttggaacacattctccctttatggtatccattaaagctctcttgaaaa 245  
Db 199 cactttacccttggaacacattctccctttatggtatccattaaagctctcttgaaaa 258  
QY 246 cctctggcattctgttgagcacctctgtgaggggctaaaggaagtgttaaatgagctgg 305  
Db 259 cctctggcattctgttgagcacctctgtgaggggctaaaggaagtgttaaatgagctgg 318  
QY 306 gaccagagctctgaagctgtgaagaaactgctgagggcgtatcacacttgggtgac 365  
Db 319 gaccagagctctgaagctgtgaagaaactgctgagggcgtatcacacttgggtgac 378  
QY 366 atcaagataaagcgagggtggtgaggaagtgccttactctctccctgcct 425  
Db 379 atcaagataaagcgagggtggtgaggaagtgccttactctctccctgcct 438  
QY 426 gaaacctgttctaccattatagatacaaatgccttaaaatgtagtgaacogtgaaaagga 485  
Db 439 gaaacctgttctaccattatagatacaaatgccttaaaatgtagtgaacogtgaaaagga 498  
QY 486 caaataaagcaatgaatacat 506  
Db 499 caaataaagcaatgaatacct 519

RESULT 6  
AAX99425  
ID AAX99425 standard; DNA; 519 BP.  
XX  
AC AAX99425;  
XX  
DT 19-OCT-1999 (first entry)  
XX  
DE iL103 specific polynucleotide.  
XX  
KW iL103; tumour; lung cancer; detection; FISH;  
KW Fluorescent in situ hybridisation; ss.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN US5939265-A.  
XX  
PD 17-AUG-1999.  
XX  
PF 05-NOV-1997; 97US-0964725.  
XX  
PR 05-NOV-1997; 97US-0964725.  
XX  
PR 05-NOV-1996; 96US-0744211.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
XX WPI; 1999-468402/39..  
XX  
XX Polynucleotides useful for detecting, diagnosing and monitoring  
PT diseases of the lung such as lung cancer  
PT  
XX  
PS Claim 1; Column 49; 36pp; English.  
XX  
XX The 519 base pair sequence of a iL103 specific polynucleotide.  
CC The polynucleotides and methods disclosed in the invention can be  
CC useful for detecting, diagnosing, staging, monitoring or predicting  
CC diseases and conditions of the lung, such as lung cancer.  
CC The polynucleotides may be used to produce probes for use in  
CC fluorescent in situ hybridization (FISH) technology to perform  
CC chromosomal analysis and identify cancer specific alterations  
CC such as deletions.  
XX

SQ Sequence 519 BP; 140 A; 113 C; 117 G; 149 T; 0 other;  
Query Match 98.5%; Score 499.4; DB 20; Length 519;  
Best Local Similarity 99.8%; Pred. No. 2.1e-153;  
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 6 gtggaacacactggtggtgattttgctagatttttctgatttttaaacctcctgaaaa 65  
Db 19 gtggaacacactggtggtgattttgctagatttttctgatttttaaacctcctgaaaa 78  
QY 66 tatccacagataaactgtoatgaagctggttaactatctctctgctggtgaccatcagccttt 125  
Db 79 tatccacagataaactgtoatgaagctggttaactatctctctgctggtgaccatcagccttt 138  
QY 126 gtagtctactctgctactgcttctctcatcaacaagtgcccttctctgttgacaagtgg 185  
Db 139 gtgttactctgctactgcttctctcatcaacaagtgcccttctctgttgacaagtgg 198  
QY 186 cactttactctctggacaacattcttctctttatggatccattaaagctcttcttgaaaa 245  
Db 199 cactttactctctggacaacattcttctctttatggatccattaaagctcttcttgaaaa 258  
QY 246 ctctgggcattctgttgagcacctctgtgaggggctaaaggaagtgtgtaaatgagctgg 305  
Db 259 ctctgggcattctgttgagcacctctgtgaggggctaaaggaagtgtgtaaatgagctgg 318  
QY 306 gaccagagctctctgaagctgtgaagaaactgctgagggcgtatcacacttgggtgac 365  
Db 319 gaccagagctctctgaagctgtgaagaaactgctgagggcgtatcacacttgggtgac 378  
QY 366 atcaagataaagcgagggtggtgaggaagtgccttactctctctccctgcct 425  
Db 379 atcaagataaagcgagggtggtgaggaagtgccttactctctccctgcct 438  
QY 426 gaaacctgttctaccattatagatacaaatgccttaaaatgtagtgaacogtgaaaagga 485  
Db 439 gaaacctgttctaccattatagatacaaatgccttaaaatgtagtgaacogtgaaaagga 498  
QY 486 caaataaagcaatgaatacat 506  
Db 499 caaataaagcaatgaatacct 519

RESULT 7  
AAX86241  
ID AAX86241 standard; CDNA; 357 BP.  
XX  
AC AAX86241;  
XX  
DT 27-APR-1999 (first entry)  
XX  
DE EST clone AA246.  
XX  
XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
XX  
OS Homo sapiens.  
XX  
PN W09845435-A2.  
XX  
PD 15-OCT-1998.  
XX  
PF 10-APR-1998; 98WO-US06954.  
XX  
PR 10-APR-1997; 97US-0835913.  
XX  
PA (GEMY ) GENETICS INST INC.  
XX  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;



```
XX US5939265-A.
XX PN
XX PD
XX PF 17-AUG-1999.
XX PF 05-NOV-1997; 97US-0964725.
XX PR 05-NOV-1997; 97US-0964725.
XX PR 05-NOV-1996; 96US-0744211.
XX PA (ABBO ) ABBOTT LAB.
XX PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX PI Kratochwill JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1999-468402/39.
XX PT Polynucleotides useful for detecting, diagnosing and monitoring
XX PT diseases of the lung such as lung cancer
XX PS Claim 1; Column 47; 36pp; English.
XX CC The 263 base pair sequence of a LU103 specific polynucleotide was
XX CC derived from clone 123531.
XX CC The polynucleotides and methods disclosed in the invention can be
XX CC useful for detecting, diagnosing, staging, monitoring or predicting
XX CC diseases and conditions of the lung, such as lung cancer.
XX CC The polynucleotides may be used to produce probes for use in
XX CC fluorescent in situ hybridization (FISH) technology to perform
XX CC chromosomal analysis and identify cancer specific alterations
XX CC such as deletions.
XX SQ Sequence 263 BP; 64 A; 57 C; 71 G; 71 T; 0 other;

Query Match 51.9%; Score 263; DB 20; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 136 tgcactgccttcctcatcaacaaagtgccttcctgttgacaaagtggcacctttacc 195
Db 1 tgcactgccttcctcatcaacaaagtgccttcctgttgacaaagtggcacctttacc 60
QY 196 tctggacaacattctccctttatggatccattaaagcttctctgaaaactctgggcat 255
Db 61 tctggacaacattctccctttatggatccattaaagcttctctgaaaactctgggcat 120
QY 256 tctgttgagcacctgtgaggggctaaagagtggtgtaaatgagctgggaccagggc 315
Db 121 tctgttgagcacctgtgaggggctaaagagtggtgtaaatgagctgggaccagggc 180
QY 316 tctgaagctgtgaaagaaactgtgagcgctatcacacttgggtgacatcaagataa 375
Db 181 tctgaagctgtgaaagaaactgtgagcgctatcacacttgggtgacatcaagataa 240
QY 376 agagcgaggtggatggggatgg 398
Db 241 agagcgaggtggatggggatgg 263

RESULT 10
AAV38066
ID AAV38066 standard; cDNA; 269 BP.
XX
XX AAV38066;
XX
XX 14-SEP-1998 (first entry)
XX
XX Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:1.
XX
XX Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.
XX
XX Homo sapiens.
```

```
XX WO9820143-A1.
XX PN
XX PD 14-MAY-1998.
XX PF 05-NOV-1997; 97WO-US20680.
XX PR 05-NOV-1996; 96US-0744211.
XX PA (ABBO ) ABBOTT LAB.
XX PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX PI Kratochwill JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1998-286957/25.
XX PT Lung tissue derived polynucleotide LU103 - useful to detect,
XX PT diagnose, stage, monitor, prognosis, prevent, treat or determine
XX PT pre-disposition to lung disease, e.g. lung cancer
XX PS Claim 1; Page 66-67; 86pp; English.
XX CC The present sequence represents a polynucleotide specific for lung
XX CC tissue gene LU103. A method has been developed for detecting the
XX CC presence of a target LU103 polynucleotide in a test sample, comprising:
XX CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,
XX CC and (b) detecting the target LU103 polynucleotide in the test sample,
XX CC where the LU103 polynucleotide has at least 50% identity to the 269,
XX CC 283, 225, 507 or 519 bp nucleic acid sequence given in AAV38086 to
XX CC AAV38070. The methods and products of the present invention may be used
XX CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or
XX CC determine the predisposition diseases and conditions of the lung, e.g.
XX CC lung cancer.
XX SQ Sequence 269 BP; 62 A; 67 C; 48 G; 92 T; 0 other;

Query Match 51.9%; Score 263; DB 19; Length 269;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 gtggaacacactggctgtggtgatttctagattttctgtatttttaaacctcctgaaaaa 65
Db 7 gtggaacacactggctgtggtgatttctagattttctgtatttttaaacctcctgaaaaa 66
QY 66 tatccagataactgctgaagctggttaactatctctcctgctgtgacacatcagccttt 125
Db 67 tatccagataactgctgaagctggttaactatctctcctgctgtgacacatcagccttt 126
QY 126 gtatttactctgctactgccttctctcatcaacaaagtgccttcctgttgacaagtgg 185
Db 127 gtatttactctgctactgccttctctcatcaacaaagtgccttcctgttgacaagtgg 186
QY 186 cacctttacactctggacaacattcttccctttatggatccattaaagcttcttctgaaaa 245
Db 187 cacctttacactctggacaacattcttccctttatggatccattaaagcttcttctgaaaa 246
QY 246 ctctgggcatctctgttgagcac 268
Db 247 ctctgggcatctctgttgagcac 269

RESULT 11
AAV99421
ID AAV99421 standard; DNA; 269 BP.
XX
XX AAV99421;
XX
XX 19-OCT-1999 (first entry)
XX
XX LU103 specific polynucleotide isolated from clone 1235095.
XX
XX LU103; tumour; lung cancer; detection; FISH;
```

```

XX Fluorescent in situ hybridisation; ss.
XX Synthetic.
OS Homo sapiens.
XX US5939265-A..
XX 17-AUG-1999.
XX
XX 05-NOV-1997; 97US-0964725.
XX
XX 05-NOV-1997; 97US-0964725.
XX
XX 05-NOV-1996; 96US-0744211.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1999-468402/39.
XX
XX Polynucleotides useful for detecting, diagnosing and monitoring
XX diseases of the lung such as lung cancer
XX
XX Claim 1; Column 47-48; 36pp; English.
XX
XX The 269 base pair sequence of a LU103 specific polynucleotide was
XX derived from clone 1235095.
XX The polynucleotides and methods disclosed in the invention can be
XX useful for detecting, diagnosing, staging, monitoring or predicting
XX diseases and conditions of the lung, such as lung cancer.
XX The polynucleotides may be used to produce probes for use in
XX fluorescent in situ hybridization (FISH) technology to perform
XX chromosomal analysis and identify cancer specific alterations
XX such as deletions.
XX
XX Sequence 269 BP; 62 A; 67 C; 48 G; 92 T; 0 other;

Query Match      51.9%; Score 263; DB 20; Length 269;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gtggaacacactgctgtggtgatttgcctagattttctgatttttaaaactctgaaaaa 65
DB 7 gtggaacacactgctgtggtgatttgcctagattttctgatttttaaaactctgaaaaa 66
QY 66 tatccagataaactgtcatgaagctgtgaactatcttctgctgtggtgaccatcagccttt 125
DB 67 tatccagataaactgtcatgaagctgtgaactatcttctgctgtggtgaccatcagccttt 126
QY 126 gtatttactctgtactgcttctctcatcaacaagaagtgcccttctctgttgacaagtgg 185
DB 127 gtatttactctgtactgcttctctcatcaacaagaagtgcccttctctgttgacaagtgg 186
QY 186 cactttactctgtgacaacattctccctttatgatccattaaagctttcttgaaaaa 245
DB 187 cactttactctgtgacaacattctccctttatgatccattaaagctttcttgaaaaa 246
QY 246 ctctgggcatctctgttgagcac 268
DB 247 ctctgggcatctctgttgagcac 269

RESULT 12
AAV38068
ID AAV38068 standard; cDNA; 225 BP.
XX
XX AAV38068;
XX
XX 14-SEP-1998 (first entry)
XX
XX Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:3.

```

```

XX Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.
XX Homo sapiens.
XX WO9820143-A1.
XX 14-MAY-1998.
XX
XX 05-NOV-1997; 97WO-US20680.
XX
XX 05-NOV-1996; 96US-0744211.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1998-286957/25.
XX
XX Lung tissue derived polynucleotide LU103 - useful to detect,
XX diagnose, stage, monitor, prognosis, prevent, treat or determine
XX pre-disposition to lung disease, e.g. lung cancer
XX
XX Claim 1; Page 67; 86pp; English.
XX
XX The present sequence represents a polynucleotide specific for lung
XX tissue gene LU103. A method has been developed for detecting the
XX presence of a target LU103 polynucleotide in a test sample, comprising:
XX (a) contacting the sample with at least 1 LU103-specific polynucleotide,
XX and (b) detecting the target LU103 polynucleotide in the test sample,
XX where the LU103 polynucleotide has at least 50% identity to the 289,
XX 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to
XX AAV38070. The methods and products of the present invention may be used
XX to detect, diagnose, stage, monitor, prognosis, prevent, treat or
XX determine the predisposition diseases and conditions of the lung, e.g.
XX lung cancer.
XX
XX Sequence 225 BP; 74 A; 41 C; 60 G; 50 T; 0 other;

Query Match      44.4%; Score 225; DB 19; Length 225;
Best Local Similarity 100.0%; Pred. No. 7.9e-64;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 aaggaagtgtgtaaatgagctgggaccagaggtcttctgaagctgtaagaaactgctgga 342
DB 1 aaggaagtgtgtaaatgagctgggaccagaggtcttctgaagctgtaagaaactgctgga 60
QY 343 ggcgtatcacacttgggtgtgacatcaagataaagcggaggtggatggggtgaaaga 402
DB 61 ggcgtatcacacttgggtgtgacatcaagataaagcggaggtggatggggtgaaaga 120
QY 403 tgatgctctctatctctcctgctgaaacctgttctaccattatagatacaatgccttaa 462
DB 121 tgatgctctctatctcctgctgaaacctgttctaccattatagatacaatgccttaa 180
QY 463 aatgtagtagccctgtaaaaggacaaataaagcaatgaatacatt 507
DB 181 aatgtagtagccctgtaaaaggacaaataaagcaatgaatacatt 225

RESULT 13
AAV99423
ID AAV99423 standard; DNA; 225 BP.
XX
XX AAV99423;
XX
XX 19-OCT-1999 (first entry)
XX
XX LU103 specific polynucleotide isolated from clone 1379417.
XX
XX LU103; tumour; lung cancer; detection; FISH;

```

KW Fluorescent in situ hybridisation; ss.  
XX Synthetic.  
OS Homo sapiens.  
XX  
XX  
PN US5939265-A.  
XX  
PD 17-AUG-1999.  
XX  
XX 05-NOV-1997; 97US-0964725.  
XX  
PF 05-NOV-1997; 97US-0964725.  
XX  
PR 05-NOV-1996; 96US-0744211.  
XX  
XX (ABBO ) ABBOTT LAB.  
XX  
XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;  
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;  
XX WPI; 1999-468402/39.  
XX  
XX Polynucleotides useful for detecting, diagnosing and monitoring  
PT diseases of the lung such as lung cancer  
XX  
XX Claim 1; Column 47; 36pp; English.  
XX  
XX The 225 base pair sequence of a LU103 specific polynucleotide was  
CC derived from clone 1379417.  
CC  
CC The polynucleotides and methods disclosed in the invention can be  
CC useful for detecting, diagnosing, staging, monitoring or predicting  
CC diseases and conditions of the lung, such as lung cancer.  
CC The polynucleotides may be used to produce probes for use in  
CC fluorescent in situ hybridization (FISH) technology to perform  
CC chromosomal analysis and identify cancer specific alterations  
CC such as deletions.  
XX  
XX Sequence 225 BP; 74 A; 41 C; 60 G; 50 T; 0 other;  
SQ

Query Match 44.4%; Score 225; DB 20; Length 225;  
Best Local Similarity 100.0%; Pred. No. 7.9e-64;  
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 aaggagtgtaaatgagctggagcagaggcttctgaagctgtgaagaactgctgga 342  
|||||  
Db 1 aaggagtgtaaatgagctggagcagaggcttctgaagctgtgaagaactgctgga 60  
|||||  
QY 343 ggcgctatcacacttggtgtgacatcaagaataaagcggagtgatggggtgaaga 402  
|||||  
Db 61 ggcgctatcacacttggtgtgacatcaagaataaagcggagtgatggggtgaaga 120  
|||||  
QY 403 tgatgctctatctccctgcctggaacctgttctccaattatagatcaaatgccctaa 462  
|||||  
Db 121 tgatgctctatctccctgcctggaacctgttctccaattatagatcaaatgccctaa 180  
|||||  
QY 463 aatgtagtgaccctgtaaaaggacaaataaagcaatgaatacatt 507  
|||||  
Db 181 aatgtagtgaccctgtaaaaggacaaataaagcaatgaatacatt 225  
|||||

RESULT 14  
AAT22859  
ID AAT22859 standard; cDNA to mRNA; 60 BP.  
XX  
XX AAT22859;  
XX  
XX 29-AUG-1996 (first entry)  
XX  
XX Human gene signature HUMGS04535.  
XX  
XX  
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.

XX Homo sapiens.  
XX  
XX WO9514772-A1.  
PN  
XX  
XX 01-JUN-1995.  
PD  
XX  
XX 11-NOV-1994; 94WO-JP01916.  
PF  
XX  
XX 12-NOV-1993; 93JP-0355504.  
PR  
XX  
XX (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
XX  
XX Matsubara K, Okubo K;  
PI  
XX WPI; 1995-206931/27.  
DR  
XX  
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
XX Claim 1; Page 1229; 2245pp; Japanese.  
PS  
XX  
XX A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AAT19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
XX  
XX Sequence 60 BP; 29 A; 10 C; 11 G; 10 T; 0 other;  
SQ

Query Match 11.4%; Score 58; DB 16; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.7e-09;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 448 gatcaaatgccctaaaaatgtagtgacccgtgaaaaggacaaataaagcaatgaataca 505  
|||||  
Db 1 gatcaaatgccctaaaaatgtagtgacccgtgaaaaggacaaataaagcaatgaataca 58  
|||||

RESULT 15  
AAV54618  
ID AAV54618 standard; cDNA; 225 BP.  
XX  
XX AAV54618;  
XX  
XX 30-OCT-1998 (first entry)  
XX  
XX LU105 specific polynucleotide sequence from clone 1605935.  
DE  
XX  
XX LU105; lung disease marker; immunoassay; lung disease; cancer;  
KW blood; plasma; serum; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9833926-A1.  
PN  
XX  
XX 06-AUG-1998.  
PD  
XX  
XX 30-JAN-1998; 98WO-US01766.  
PF



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 12:47:26 ; Search time 4216.79 Seconds  
(without alignments)  
1859.744 Million cell updates/sec

Title: US-09-700-770-1  
Perfect score: 507  
Sequence: 1 ggcaagtgaaccactgct.....aataaagcaatgaatacatt 507

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_ba3:  
4: gb\_in1:  
5: gb\_in2:  
6: gb\_in3:  
7: gb\_om:  
8: gb\_ov:  
9: gb\_pat1:  
10: gb\_pat2:  
11: gb\_ph:  
12: gb\_pl1:  
13: gb\_pl2:  
14: gb\_pl3:  
15: gb\_pl4:  
16: em\_ba1:  
17: em\_ba2:  
18: em\_fun:  
19: em\_htgo\_hum:  
20: em\_htgo\_inv:  
21: em\_htgo\_rod:  
22: em\_htg\_hum1:  
23: em\_htg\_hum2:  
24: em\_htg\_hum3:  
25: em\_htg\_hum4:  
26: em\_htg\_hum5:  
27: em\_htg\_hum6:  
28: em\_htg\_hum7:  
29: em\_htg\_hum8:  
30: em\_htg\_inv1:  
31: em\_htg\_inv2:  
32: em\_htg\_other:  
33: em\_htg\_rod:  
34: em\_hum1:  
35: em\_hum2:  
36: em\_hum3:  
37: em\_hum4:  
38: em\_hum5:  
39: em\_hum6:  
40: em\_hum7:  
41: em\_in:  
42: em\_om:  
43: em\_or:

Result No.	Score	Query Match	Length	DB ID	Description
1	205.4	40.5	166539	62	AC011402 Homo sapi
2	192.8	38.0	160949	62	AC011352 Homo sapi
3	157	31.0	207747	62	AC011334 Homo sapi
C 4	57.8	11.4	96482	67	AC022095 Homo sapi
C 5	57.8	11.4	168347	69	AC025336 Homo sapi
6	56.4	11.1	471	97	HUMZB52D10
7	38.2	7.5	148391	71	AC034174
8	38	7.5	7218	10	I66494 Sequence 14

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.







*****		Summary Statistics	
Consensus quality: 73731 bases at least Q40		Consensus quality: 84145 bases at least Q30	
Consensus quality: 84145 bases at least Q30		Consensus quality: 88496 bases at least Q20	
Estimated insert size: 131000; pulse field gel estimation		Estimated insert size: 93182; sum-of-contigs estimation	
Quality coverage: 4.03 in Q20 bases; pulse field gel estimation		Quality coverage: 5.54 in Q20 bases; sum-of-contigs estimation.	
* NOTE: This is a 'working draft' sequence. It currently		* consists of 14 contigs. The true order of the pieces	
* is not known and their order in this sequence record is		* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.		* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will		* be preserved.	
1		2069: contig of 2069 bp in length	
2070		2169: gap of unknown length	
2170		4188: contig of 2019 bp in length	
4189		4288: gap of unknown length	
4289		6338: contig of 2050 bp in length	
6439		6438: gap of unknown length	
6439		8789: contig of 2351 bp in length	
8790		8889: gap of unknown length	
8890		11124: contig of 2235 bp in length	
11125		11224: gap of unknown length	
13473		13472: contig of 2248 bp in length	
13573		13572: gap of unknown length	
16710		16709: contig of 3137 bp in length	
16810		16809: gap of unknown length	
20135		20134: contig of 3325 bp in length	
20235		20234: gap of unknown length	
23845		23845: contig of 3611 bp in length	
23946		23945: gap of unknown length	
28527		28527: contig of 4582 bp in length	
28628		28627: gap of unknown length	
34285		34284: contig of 5657 bp in length	
34385		34384: gap of unknown length	
45173		45172: contig of 10788 bp in length	
45273		45272: gap of unknown length	
56034		56033: contig of 10761 bp in length	
56134		56133: gap of unknown length	
96482		96482: contig of 40349 bp in length.	
FEATURES		Location/Qualifiers	
1..96482		/organism="Homo sapiens"	
/db_xref="taxon:9606"		/chromosome="5"	
/clone_lib="Caltech human BAC library B"		/clone="CTB-36B8"	
BASE COUNT		21454 a 25145 c 26218 g 22348 t 1317 others	
ORIGIN			
Query Match		11.4%; Score 57.8; DB 67; Length 96482;	
Best Local Similarity		59.4%; Pred. No. 2.4e-05;	
Matches		98; Conservative 0; Mismatches 67; Indels 0; Gaps	
Qy		195 ctcctggacaacattcttcctttatgatccattaaagcttctctgaaacctgggcca 254	
Db		18605 CCCTGGCCCAACCCCTGGCACCCTCAACCGCTGAAGCTCTGCTGACGAGCCTGGGCA 18546	
Qy		255 tttctgttgagcactgttgaggagggttaaggaagtgtgtaaatgactgggacaagg 314	
Db		18545 TCCCGGTGAACCACTCATATGAGGGCTCCACAGAGTGTGGCTGGTGGTCCCCCAGG 18486	
Qy		315 cttctggaagtgtgaagaaactgctggaggcgctatcacacttgg 359	
Db		18485 CCGTGGGGCGGTGAAGGCCCTGAAGGCCCTGCTGGTAAAGTGGG 18441	
RESULT		5	

```

AC025336/c
LOCUS       AC025336       168347 bp       DNA             HTG             25-MAR-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-451H23 map 5, WORKING DRAFT
SEQUENCE, 32 unordered pieces.
ACCESSION   AC025336
VERSION     AC025336.2   GI:7328761
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 168347)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 5, clone RP11-451H23
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 168347)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhagter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Headford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczeky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
McElrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 25, 2000 this sequence version replaced gi:7210017.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6586
Center clone name: 451_H_23
----- Summary Statistics
Sequencing vector: M13, M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150422 bases at least Q40
Consensus quality: 159524 bases at least Q30
Consensus quality: 163013 bases at least Q20
Insert size: 165247; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1389: contig of 1389 bp in length
* 1390 1489: gap of 100 bp
* 1490 3130: contig of 1641 bp in length

```

## FEATURES

Source

```

1..168347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"
/clone_lib="RP11-451H23"
/clone="RP11-451H23"
1..1389
/note="assembly_fragment"
misc_feature
1490..3130
/note="assembly_fragment"
misc_feature
3231..4942

```

```
/note="assembly_fragment"
5043. .6981
/note="assembly_fragment"
7082. .8708
/note="assembly_fragment"
8809. .10286
/note="assembly_fragment"
10387. .12212
/note="assembly_fragment"
12313. .14658
/note="assembly_fragment"
14759. .17941
/note="assembly_fragment"
18042. .21297
/note="assembly_fragment"
21398. .24992
/note="assembly_fragment"
25093. .27768
/note="assembly_fragment"
27869. .31188
/note="assembly_fragment"
31289. .33714
/note="assembly_fragment"
33815. .37277
/note="assembly_fragment"
37378. .42302
/note="assembly_fragment"
42403. .47816
/note="assembly_fragment"
47917. .52586
/note="assembly_fragment"
52687. .56567
/note="assembly_fragment"
56668. .61557
/note="assembly_fragment"
61658. .66724
/note="assembly_fragment"
66825. .71568
/note="assembly_fragment"
71669. .76578
/note="assembly_fragment"
76679. .83312
/note="assembly_fragment"
83413. .90053
/note="assembly_fragment"
90154. .99426
/note="assembly_fragment"
99527. .108015
/note="assembly_fragment"
108116. .118144
/note="assembly_fragment"
118245. .130468
/note="assembly_fragment"
130569. .142239
/note="assembly_fragment"

Query Match      11.4%; Score 57.8; DB 69; Length 168347;
Best Local Similarity 59.4%; Pred. No. 2.7e-05;
Matches 98; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 195 ccttgacaacattcttcccttttgatgcatcattaaagcttcttctgaaactctgggca 254
Db 143960 CCCGTGGCCAAACCCCTCGGGCAACCCCTCAACCCGCTGAAGCTCCTGCTGAGCAGCCTGGGCA 143901

QY 255 ttctgttgagcaccttgtgagggttaaggaagtgtgtaaatgagctgggaccagag 314
Db 143900 TCCCGTGAACACCTCATAGAGGGCTCCAGAGAGTGTGTGGCTGAGCTGGGTCCCGAGG 143841

QY 315 ctctgaagctgtgaagaaactgtgaggcgctatcacacttgg 359
Db 143840 CCGTGGGGCCGTGAAGCCCTGAAGCCCTGCTGGTGAAGTGGG 143796

/note="assembly_fragment"
5043. .6981
/note="assembly_fragment"
7082. .8708
/note="assembly_fragment"
8809. .10286
/note="assembly_fragment"
10387. .12212
/note="assembly_fragment"
12313. .14658
/note="assembly_fragment"
14759. .17941
/note="assembly_fragment"
18042. .21297
/note="assembly_fragment"
21398. .24992
/note="assembly_fragment"
25093. .27768
/note="assembly_fragment"
27869. .31188
/note="assembly_fragment"
31289. .33714
/note="assembly_fragment"
33815. .37277
/note="assembly_fragment"
37378. .42302
/note="assembly_fragment"
42403. .47816
/note="assembly_fragment"
47917. .52586
/note="assembly_fragment"
52687. .56567
/note="assembly_fragment"
56668. .61557
/note="assembly_fragment"
61658. .66724
/note="assembly_fragment"
66825. .71568
/note="assembly_fragment"
71669. .76578
/note="assembly_fragment"
76679. .83312
/note="assembly_fragment"
83413. .90053
/note="assembly_fragment"
90154. .99426
/note="assembly_fragment"
99527. .108015
/note="assembly_fragment"
108116. .118144
/note="assembly_fragment"
118245. .130468
/note="assembly_fragment"
130569. .142239
/note="assembly_fragment"

Query Match      11.4%; Score 57.8; DB 69; Length 168347;
Best Local Similarity 59.4%; Pred. No. 2.7e-05;
Matches 98; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 195 ccttgacaacattcttcccttttgatgcatcattaaagcttcttctgaaactctgggca 254
Db 143960 CCCGTGGCCAAACCCCTCGGGCAACCCCTCAACCCGCTGAAGCTCCTGCTGAGCAGCCTGGGCA 143901

QY 255 ttctgttgagcaccttgtgagggttaaggaagtgtgtaaatgagctgggaccagag 314
Db 143900 TCCCGTGAACACCTCATAGAGGGCTCCAGAGAGTGTGTGGCTGAGCTGGGTCCCGAGG 143841

QY 315 ctctgaagctgtgaagaaactgtgaggcgctatcacacttgg 359
Db 143840 CCGTGGGGCCGTGAAGCCCTGAAGCCCTGCTGGTGAAGTGGG 143796

/note="assembly_fragment"
5043. .6981
/note="assembly_fragment"
7082. .8708
/note="assembly_fragment"
8809. .10286
/note="assembly_fragment"
10387. .12212
/note="assembly_fragment"
12313. .14658
/note="assembly_fragment"
14759. .17941
/note="assembly_fragment"
18042. .21297
/note="assembly_fragment"
21398. .24992
/note="assembly_fragment"
25093. .27768
/note="assembly_fragment"
27869. .31188
/note="assembly_fragment"
31289. .33714
/note="assembly_fragment"
33815. .37277
/note="assembly_fragment"
37378. .42302
/note="assembly_fragment"
42403. .47816
/note="assembly_fragment"
47917. .52586
/note="assembly_fragment"
52687. .56567
/note="assembly_fragment"
56668. .61557
/note="assembly_fragment"
61658. .66724
/note="assembly_fragment"
66825. .71568
/note="assembly_fragment"
71669. .76578
/note="assembly_fragment"
76679. .83312
/note="assembly_fragment"
83413. .90053
/note="assembly_fragment"
90154. .99426
/note="assembly_fragment"
99527. .108015
/note="assembly_fragment"
108116. .118144
/note="assembly_fragment"
118245. .130468
/note="assembly_fragment"
130569. .142239
/note="assembly_fragment"

Query Match      11.4%; Score 57.8; DB 69; Length 168347;
Best Local Similarity 59.4%; Pred. No. 2.7e-05;
Matches 98; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 195 ccttgacaacattcttcccttttgatgcatcattaaagcttcttctgaaactctgggca 254
Db 143960 CCCGTGGCCAAACCCCTCGGGCAACCCCTCAACCCGCTGAAGCTCCTGCTGAGCAGCCTGGGCA 143901

QY 255 ttctgttgagcaccttgtgagggttaaggaagtgtgtaaatgagctgggaccagag 314
Db 143900 TCCCGTGAACACCTCATAGAGGGCTCCAGAGAGTGTGTGGCTGAGCTGGGTCCCGAGG 143841

QY 315 ctctgaagctgtgaagaaactgtgaggcgctatcacacttgg 359
Db 143840 CCGTGGGGCCGTGAAGCCCTGAAGCCCTGCTGGTGAAGTGGG 143796
```

RESULT 6  
HUMZB52D10  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

HUMZB52D10 471 bp mRNA PRI 29-AUG-1998  
Homo sapiens full length insert cDNA clone ZB52D10.  
AF086152  
AF086152.1 GI:3483497  
FLI\_CDNA.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 471)  
Woessner,J., Tan,F., Marra,M., Kucaba,T., Vandell,M., Martin,J.,  
Marth,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,  
Geisel,S., Allen,M., Underwood,K., Chappell,J., Persson,B.,  
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,  
Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behymer,K., Hillier,L.,  
Willson,R. and Waterston,R.  
Full Clone Sequencing of the Longest Available Member from Each  
Unigene Cluster  
Unpublished  
2 (bases 1 to 471)  
Waterston,R.  
Direct Submission  
Submitted (24-AUG-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
SUBMITTED BY:  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No  
attempt has been made to verify whether this corresponds to the  
full-length of the original mRNA from which it was derived. We  
have tried to obtain double-stranded, or double chemistry sequence  
across the entire clone, but potentially, there are areas in the  
sequence where this level of coverage was not achieved.  
Nevertheless, we are confident of the accuracy of this sequence as  
all regions of low quality, as defined by PHRAP (P. Green, in  
preparation), were visually inspected and edited accordingly. The  
consensus quality values for this sequence have been submitted  
separately.

The location of this clone is unknown.

FEATURES  
Source

Location/Qualifiers  
1..471  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares\_fetal\_lung\_NBHL19W"  
/clone="IMAGE:307219"  
BASE COUNT 78 a 171 c 149 g 73 t  
ORIGIN

Query Match 11.1%; Score 56.4; DB 97; Length 471;  
Best Local Similarity 60.4%; Pred. No. 2.6e-05;  
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 195 ccttgacaacattcttcccttttgatgcatcattaaagcttcttctgaaactctgggca 254  
Db 158 CCGTGGCCAAACCCCTCGGGCAACCCCTCAACCCGCTGAAGCTCCTGCTGAGCAGCCTGGGCA 217  
QY 255 ttctgttgagcaccttgtgagggttaaggaagtgtgtaaatgagctgggaccagag 314  
Db 218 TCCCGTGAACACCTCATAGAGGGCTCCAGAGAGTGTGTGGCTGAGCTGGGTCCCGAGG 314  
QY 315 ctctgaagctgtgaagaaactgtgaggcgct 348  
Db 278 CCGTGGGGCCGTGAAGCCCTGAAGCCCTGCTGGTGAAGTGGG 311



```

RESULT 8
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheifflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 7.5%; Score 38; DB 10; Length 7218;
Best Local Similarity 4.3%; Pred. No. 6.2;
Matches 11; Conservative 144; Mismatches 99; Indels 0; Gaps 0;

Qy 18 gcttggtgatttgcagattttctgatttttaaaactcctgaaataatccagataa 77
Db 1048 GTCGAGGAGCTGCCATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1107

Qy 78 ctgctgaagctgtaactatctctgctggtgaccatcagcccttctgagttactc 137
Db 1108 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1167

Qy 138 ctactgcttcctcaacaagaagtcctcctctgtgacagttggacattacctc 197
Db 1168 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1227

Qy 198 tggacacattctcctcttatgacataaagctctctctgaaactctggcattt 257
Db 1228 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1287

Qy 258 ctgttgagcactt 271
Db 1288 YYYYYYYYYYYYYY 1301

RESULT 9
LOCUS AL356463 154268 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-100E13, *** SEQUENCING IN
PROGRESS ***, 17 unordered pieces.
ACCESSION AL356463
VERSION AL356463.4 GI:10039903
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 154268)
Plumb, B.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:9650562.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BAL00E13
----- Summary Statistics

```

```

Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 145752 bases at least Q40
Consensus quality: 149198 bases at least Q30
Consensus quality: 151149 bases at least Q20
Insert size: 152668; sum-of-contigs
Quality coverage: 3.51x in Q20 bases; agarose-fp
coverage: 3.44x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1
* 4950 5049: gap of 100 bp in length
* 5050 8238: contig of 3189 bp in length
* 8239 8338: gap of 100 bp
* 8339 20035: contig of 11697 bp in length
* 20036 20135: gap of 100 bp
* 20136 22459: contig of 3324 bp in length
* 23460 23539: gap of 100 bp
* 23560 40503: contig of 16944 bp in length
* 40604 44208: contig of 3605 bp in length
* 44209 44308: gap of 100 bp
* 44309 47023: contig of 2715 bp in length
* 47024 47123: gap of 100 bp
* 47124 51832: contig of 4709 bp in length
* 51833 51932: gap of 100 bp
* 51933 55546: contig of 3614 bp in length
* 55547 55646: gap of 100 bp
* 55647 64021: contig of 8375 bp in length
* 64022 64121: gap of 100 bp
* 64122 68879: contig of 4758 bp in length
* 68880 68979: gap of 100 bp
* 68980 86316: contig of 17337 bp in length
* 86317 86416: gap of 100 bp
* 86417 132139: contig of 45723 bp in length
* 132140 132239: gap of 100 bp
* 132240 137720: contig of 5481 bp in length
* 137721 137820: gap of 100 bp
* 137821 141195: contig of 3375 bp in length
* 141196 141295: gap of 100 bp
* 141296 147431: contig of 6136 bp in length
* 147432 147531: gap of 100 bp
* 147532 154268: contig of 6737 bp in length.

```

```

FEATURES
Source
1..154268
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-100E13"
/clone_lib="RPC1-11.1"
1..4949
/note="assembly_fragment:00948
clone_end:SP6
vector_side:left"
5050..8238
/note="assembly_fragment:00148
fragment_chain:1"
8339..20035
/note="assembly_fragment:00915
fragment_chain:1"
20136..23459
/note="assembly_fragment:00542
fragment_chain:1"
23560..40503
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```



```

Db 59619 CAGCAAGAAAGAAATATTAT 59600
|| | ||| || |
ALL138760 172026 bp DNA HTG 25-FEB-2001
Homo sapiens chromosome 10 clone RP11-123B3, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ALL138760
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sims,S.
Direct Submission
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:13160171.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA123B3
----- Summary Statistics
Sequencing vector: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 17118 bases at least Q40
Consensus quality: 171427 bases at least Q30
Consensus quality: 171646 bases at least Q20
Insert size: 171926; sum-of-contigs
Insert size: 170954; 3.2% error; agarose-fp
Quality coverage: 6.64x in Q20 bases; sum-of-contigs Quality
coverage: 6.76x in Q20 bases; agarose-fp
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 116995: contig of 116995 bp in length
116996 117095: gap of 100 bp
117096 172026: contig of 54931 bp in length.
Location/Qualifiers
1..172026
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-123B3"
/clone_lib="RPC1-11.1"
1..116995
/note="assembly fragment:01291
fragment_chain:1
clone_end:SP6
vector_side:left"
117096..172026
/note="assembly fragment:00073
fragment_chain:1"
46605 a 35115 c 38947 g 51259 t 100 others
7.4%; Score 37.6; DB 78; Length 172026;

FEATURES
Source
misc_feature
misc_feature
BASE COUNT
ORIGIN
Query Match

```

```

Best Local Similarity 56.5%; Pred. No. 13;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 144 cttctcatcaacaagaagtccttctgtgacaaagttggcacctttacctctggaca 203
|| || || || || || || || || || || || || || || || || || || || || ||
Db 101462 CCCTCTTCATTCAGAAAGTTCTCCTTTCTATAGCCCATCAGCTGTGTGCAGCTCTTAT 101521
|| || || || || || || || || || || || || || || || || || || || || ||
QY 204 acattcttcccttatgatccattataaagcttcttctgaaaacctctggcattctctgtg 263
|| || || || || || || || || || || || || || || || || || || || || ||
Db 101522 TCTTTTCTCCAGTTTGTGACCCCTCAAGGTTTTTTTAAAAATATTATATCTCTGTGA 101581
|| || || || || || || || || || || || || || || || || || || || || ||
QY 264 agca 267
|| || ||
Db 101582 TGCA 101585
|| || ||

RESULT 12
AC083856/c
LOCUS
DEFINITION
AC083856
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Gupta,J., Ho,S.-L., Idol,J.,
Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,O.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B.,
Stantropop,S., Thomas,J.W., Thomas,P.J., Tlionson,E.E.,
Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A., Wetherby,K.D. and
Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 215455)
Green,E.D.
Direct Submission
Submitted (04-OCT-2000) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
----- Project Information
Center project name: sh
Center clone name: 366H13
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 207240 bases at least Q40
Consensus quality: 209922 bases at least Q30
Consensus quality: 211032 bases at least Q20
Insert size: 214000; agarose-fp
Insert size: 229000; pulse-field-gel
Insert size: 213955; sum-of-contigs
Quality coverage: 5.63x in Q20 bases; agarose-fp
Quality coverage: 5.26x in Q20 bases; pulse-field-gel
Quality coverage: 5.63x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```





entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids. Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES  
Source

Location/Qualifiers

```

1. .185286
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="Yp"
/clone="GSMB-187H15"
join(386. .486,2580. .2641,4014. .4062,6524. .6653,7385. .7540,
7967. .8066,17122. .17257,19778. .19864,20273. .20371,
20447. .20607,20993. .21137,21247. .21391,21757. .21938,
22008. .22125,23018. .23171,23267. .23397,23600. .25372)
/genes="Mus musculus mRNA for DBY RNA helicase AJ007376"
complement(1086. .1201)
/rpt_family="B1_MM"
repeat_region
complement(1523. .1645)
/rpt_family="B3"
repeat_region
complement(2224. .2438)
/rpt_family="B3"
repeat_region
4289. .4421
/rpt_family="B1-F"
repeat_region
4617. .4785
/rpt_family="B3"
repeat_region
complement(5462. .5644)
/rpt_family="B2"
repeat_region
complement(6735. .7044)
/rpt_family="MT2B"
repeat_region
complement(7826. .7852)
/rpt_family="AT_rich"
repeat_region
complement(9772. .12427)
/rpt_family="IAP"
repeat_region
complement(12822. .13010)
/rpt_family="HERVK"
repeat_region
complement(13168. .14064)
/rpt_family="IAP"
repeat_region
complement(15303. .15357)
/rpt_family="(CCAA)n"
repeat_region
complement(15425. .15581)
/rpt_family="B2"
repeat_region
complement(15595. .16288)
/rpt_family="Lx9"

```

```

repeat_region
complement(16638. .16706)
/rpt_family="B4A"
repeat_region
complement(16695. .16871)
/rpt_family="B3"
repeat_region
complement(17360. .17412)
/rpt_family="ORR1B"
repeat_region
complement(17554. .17726)
/rpt_family="ORR1B"
17793. .17938
/rpt_family="B1_MM"
repeat_region
17960. .18000
/rpt_family="(GA)n"
repeat_region
18582. .18691
/rpt_family="(GAAA)n"
repeat_region
complement(18696. .18883)
/rpt_family="ORR1C"
19321. .19403
/rpt_family="B1-F"
repeat_region
complement(19518. .19589)
/rpt_family="PB1D9"
20065. .20208
/rpt_family="B3"
repeat_region
complement(21535. .21619)
/rpt_family="B4A"
22548. .22593
/rpt_family="B1-F"
repeat_region
22711. .22846
/rpt_family="B1-F"
repeat_region
complement(23403. .23451)
/rpt_family="(CAAA)n"
24630. .24700
/rpt_family="(TAAAA)n"
repeat_region
25239. .25387
/rpt_family="B1_MM"
repeat_region
26303. .26449
/rpt_family="B1_MM"
26534. .26705
/rpt_family="B2"
repeat_region
complement(26852. .26972)
/rpt_family="B1_MM"
repeat_region
complement(27306. .27426)
/rpt_family="B1_MM"
27452. .27653
/rpt_family="Lx8"
repeat_region
28137. .28342
/rpt_family="B2"
repeat_region
30010. .30107
/rpt_family="MYSERV"
repeat_region
complement(30147. .30346)
/rpt_family="B2"
30955. .31375
/rpt_family="RMR17C"
repeat_region
complement(31444. .31926)
/rpt_family="RMR15"
repeat_region
complement(31927. .32058)
/rpt_family="ORR1B"
repeat_region
complement(32065. .32178)
/rpt_family="RMR15"
repeat_region
complement(32885. .33207)
/rpt_family="RLTRETN_MM"
repeat_region
complement(34002. .34115)
/rpt_family="MT2B"
34450. .34552
/rpt_family="B2"
repeat_region
35167. .35356
/rpt_family="B2"
repeat_region
35526. .35626
/rpt_family="RSINEL"
repeat_region
35920. .35996
/rpt_family="B3A"
repeat_region
35997. .36095
/rpt_family="RLTRI3B"
36107. .36953

```

[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 13:33:57 ; Search time 2570.26 Seconds  
(without alignments)  
7576.229 Million cell updates/sec

Title: US-09-700-770-3  
Perfect score: 2060  
Sequence: 1 cttgagagctctcaataact.....ttccattgaaaaaaaaa 2060

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



```
/note="putative"
/codon_start=1
/protein_id="BAB22158.1"
/db_xref="GI:12832561"
/translation="MSPILLLLLLCLLLGLNLEPEAKLRVPLQRIHLGHRILNPLNGW
EQALERTSTSGGNPSFVPLSKFNTQFGTIGLTPQNPVVFDTGSSNLWVPST
RCHFSLACFWHFRNPKASFRNGTKFAIQYGTGRSLGSLSDNLEKIPVFSFLNR
TGEALMPSFLHAFHFDGILGLFPLTAGVGPPLDAMVEOGLLEKIPVFSFLNR
DESGDGLSLVGGSDPAHYVPLPFIPTVPAWQVHMSKVYGTGLSLCAQCSAI
LDGTSLITGPEEIRNALNKAIIGVYFPLNGOYFIQCSKTPILPVPSPHLSGWNPLTG
ODYVTKIQSDVGLCLLGFQALDIPNAGPLWILGDVFLGPIVAVFDRGNVGRVVG
LARAOSRSTDRRAETTTQAQFFKRRP"
263 a 384 c 356 g 328 t
BASE COUNT
ORIGIN

Query Match      38.3%; Score 789; DB 192; Length 1331;
Best Local Similarity 76.0%; Pred. No. 1.5e-207;
Matches 118; Conservative 0; Mismatches 305; Indels 16; Gaps 3;

QY 694 ajaacgcgtccacagatgtctccaccaccgcgtctgcaacccctgctgctgctgtgc 753
Db 2 AACTAGTCCAGAGATGTCGCACTACTGCTG-----CTGCTGCTGTCCTGC 52

QY 754 c-ctgctgaatgtgagccttcgagccacacactgacgcacccctcttcacatagctcc 813
Db 53 TCTGGGGAATTTGAGCCTGAGGAGGCCAACTGATCGTGTCCCTCTTCAACGAATCC 112

QY 814 auctggacgcagggccctgaacctactgaggggatgagagaaccagcagagctcccca 873
Db 113 A-CTTGGACACAGAACTCTTAACCCACTGAATGATGGAGTGGACAGTGCAGAGCTTCTTA 172

QY 874 aattggggcccatccctcctggggacagcccatcttcgactctctcgaactacaggg 933
Db 173 -----GGACCTCCACCTCTGGTGCAACCCCTCTCTTTGTGCTCTCTCCAAGTTCATGA 226

QY 934 atgtcagttatttgggaataattggctgggaacgcctcccaaaacttcactgtgct 993
Db 227 ACACCCAGTATTGGAACTATTGGTTGGGAACGCCCTCTCAGAAATTTACCCGTGTCT 286

QY 994 tgaacatgctctccaatctctgggtcccgctccagagagatgccactcttcagttgc 1053
Db 287 TGACACGGGTCTTCAACTTGTGGTTCGCTCCACAGATGTCAATTTCTTCAAGTTGG 346

QY 1054 ctgctggttacacacagcattatcccaagcctctagctctccagggcccaatggga 1113
Db 347 CATGCTGTTTCACCATCGCTTTAATCCCAAGGCTCCAGCTCTCTCAGGCCCAATGGGA 406

QY 1114 ccaagttgcatcattgaatactggcgggttagatgggaatcctgagcgagacaagc 1173
Db 407 CCAAGTTTGCCATTTCAGTATGGGACCGGGGCTGAGCGGAATCTGAGCCAGACAACTC 466

QY 1174 tgactattgtggaatacaaggtgcatcagtgatttcggggaggtctctggagccca 1233
Db 467 TGACTATCGGGGGATCCACGATGCTTTGTGACATTTGGAGAGGCTCTGTGGAGGCCA 526

QY 1234 gcctggtctctgttttgcattttgatggatattggcctcggtttcccaatttgt 1293
Db 527 GCCTGATCTTTAGCCACTTTGATGGGATCTTGGGCTCGGGTTCCTCCACTCTGG 586

QY 1294 ctgtgaagagtgctggcccccgatgatgtactggtggagcaggggctattggataaagc 1353
Db 587 CTGTGGCGGAGTTTCAGCCTCCGCTGGATGGATGGTGGGAGCAAGGCTCTCTGGAGAAC 646

QY 1354 ctgtctctctcttaccacaggaacctgaagacgtgatgagagagactggtcc 1413
Db 647 CCCTCTCTCTCTTACTCAACAGGATTTCTGAAGGCTGTGATGGGGGAGAGCTGTCTC 706

QY 1414 tgggggctcgagccggcacatacatccaccctccactctcgtgagtcacaggtcc 1473
Db 707 TAGGGGCTCAGACCCCGCTCACTACGTACTTCCCTTCACTTCACTACAGTCAACATCC 766

QY 1474 ctgcctactggcagatccacatggcgtgtgaggtggtggccagggctgactctctgtg 1533
```

```
Db 767 CTGCCTACTGGCAGGTCCACATGGAGAGTGTGAAGTGGCACAGGCTGAGCTCTGTG 826
QY 1534 ccaaggctgtgctccactcctcctgatacgggcacgtccctcatcacagaccactgag 1593
Db 827 CCACGGCTGCAGTCCATCTAGACAGGACATCCCTCATCAGAGCTTAGTAGG 886
QY 1594 agatccggcctgcagtcagccattgggggaatccctgtgctggctgggagtagatca 1653
Db 887 AGATCCGGGCTTGAATAAAGCCATTGGGGATATCCCTTCTGAATGGCAGTACTTCA 946
QY 1654 tccgtgctcggaaatcccaagctcccgagctcctctctctctctcttcttggggggtcgt 1713
Db 947 TTCAGTGTTCACAGCGCAACGCTTCCCTCTGCTCTCCACCTGGTGGAGTCTGGT 1006
QY 1714 ttaacctcagcccatgattacgtcatccagactactcgaatggcgtccgctctgct 1773
Db 1007 TTAACCTCACAGGCGAGGACTATGTCAAGATTCCTCAGAGCGATGTTGGCCTCTGCC 1066
QY 1774 tgcggtttccagccctggatgtccctcgcgtcagggcctctctgctgctcgtg 1833
Db 1067 TGTGGGCTTCCAGGCTTGGATATCCCAACGCTGCGGGACCCCTCTGATCTCTGGG 1126
QY 1834 acgtctcttgggacgtatgtgacctcttcgacctcgcagggcctctctgctcgtg 1893
Db 1127 ACCTCTTTTGGGCGCTATGTGCTGTCTTGAACCGTGGGACAGAAAGGACTACGC 1186
QY 1894 ggtggtcctggcgcgcgtcgcactcgcagcgagcgacctcgatggggagagactcgc 1953
Db 1187 GCGTGGGACTGGCGGTGCTCAGTCTGCTTCAACAGACCGCGGACAGAAAGGACTACGC 1246
QY 1954 aggcgcagttcccggtgacgcccagaagta-aagcgcagtcgcagcggtggtcgagag 2012
Db 1247 AGGCGAGTCTTCAAAAGACGCCCTGTTAGGTACAGCTCACCGGCGCCACAGCAGCT 1306
QY 2013 gtctgtaccaccagtaaa 2031
Db 1307 ATGCTTCTTCCCAATTA 1325
```

## RESULT 2

```
LOCUS BG540257 878 bp mRNA EST 03-APR-2001
DEFINITION 602569148F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4693859 5',
rna sequence.
ACCESSION BG540257
VERSION BG540257.1 GI:13532490
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 878)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Inceye Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1516 row: o column: 12
High quality sequence stop: 718.
Location/Qualifiers
1. .878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4693859"
/clone_lib="NIH_MGC_77"
```

## FEATURES source













```

High quality sequence stop: 452.
Location/Qualifiers
1. .612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:244915"
/clone_lib="NCL CGAP-Lu19"
/tissue_type="squamous cell carcinoma, poorly
metastatic"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: p773D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p773 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
128 a 187 c 197 g 100 t

BASE COUNT
ORIGIN

Query Match 29.4%; Score 605.2; DB 103; Length 612;
Best Local Similarity 99.5%; Pred. No. 1.1e-156;
Matches 507; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1440 a cccaccctcactctgtccagtcacggctccctgctactgctgagatccacatgag 1499
Db 612 a cccaccctcactctgtccagtcacggctccctgctactgctgagatccacatgag 1499
QY 1500 cgttgaagtgaggccagggctgactctctgtgcaaggggtgtgtccatctgtgat 1559
Db 552 cgttgaagtgaggccagggctgactctctgtgcaaggggtgtgtccatctgtgat 1559
QY 1560 a gggcagctccctcactacacagagaccactgagagatccggcctcatgacacatt 1619
Db 492 a gggcagctccctcactacacagagaccactgagagatccggcctcatgacacatt 1619
QY 1620 g cgggaatccctctgctgggaggtacatcctgtgctggaaatcccaagctc 1679
Db 432 g cgggaatccctctgctgggaggtacatcctgtgctggaaatcccaagctc 1679
QY 1680 c cgcagctcctctctctgtgggggtctgtgttaacctcacggcccatgattagctc 1739
Db 372 c cgcagctcctctctctgtgggggtctgtgttaacctcacggcccatgattagctc 1739
QY 1740 a tccagactactcgaatggtcgcgcctctgtgtccggtttccagggccttgatgtc 1799
Db 312 a tccagactactcgaatggtcgcgcctctgtgtccggtttccagggccttgatgtc 1799
QY 1800 c tccgcctgcagggccctctgtgatctcgtgaactcttcttgggagcgtatgtggcc 1859
Db 252 c tccgcctgcagggccctctgtgatctcgtgaactcttcttgggagcgtatgtggcc 1859
QY 1860 g ttttcgaccgcgggacatgaagcagcgcgcgggtgggcctggcgcgtcgcact 1919
Db 192 g ttttcgaccgcgggacatgaagcagcgcgcgggtgggcctggcgcgtcgcact 1919
QY 1920 c gtagcagacactcgtatgggagagactcgcagcgcagttcccccggtgacgcca 1979
Db 132 c gtagcagacactcgtatgggagagactcgcagcgcagttcccccggtgacgcca 1979
QY 1980 a gtagcagacactcgtatgggagagactcgcagcgcagttcccccggtgacgcca 1979
Db 72 a gtagcagacactcgtatgggagagactcgcagcgcagttcccccggtgacgcca 1979
QY 2040 t tccattga 2049
Db 12 t tccattga 2049

```

```

RESULT 10
AL545162
LOCUS
DEFINITION
AL545162 LTI_NFL006.PL2 Homo sapiens cDNA clone CSOD1028YB18 5
prime, mRNA sequence.
ACCESSION
AL545162
VERSION
AL545162.1 GI:12877643
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 933)
AUTHORS
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSOD1028YB18"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/notes="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 278 a 191 c 219 g 244 t 1 others
ORIGIN

```

```

Query Match 28.5%; Score 587.4; DB 106; Length 933;
Best Local Similarity 99.8%; Pred. No. 1.1e-151;
Matches 588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 117 tgaccaagaaggttcaaaaacttcagcagcagctctgaagaatcgtgaaatgtgcg 176
Db 1 TGACCAAGAAGGTTCAAAAACCTTCAGCGAGCAGCTCTGAAGAATCTGTGAAATGTGCGG 60
QY 177 tttctctaaataccagacagcttgaaaaattacagcaattattatttttccctcta 236
Db 61 TTTCTCTAAATACCAAGAGTTGAAAAATTACAGCAATATTATATTTTATTCCTCTA 120
QY 237 aattcaagatcacctcctgggttatctctaaatgaattggctggaaactccttatga 296
Db 121 AATCAAGGATACCTACCTGGTATTATTTCTAAATGAATTGGCTGGAATCTCTTATGA 180
QY 297 tatctgcagcacttgtaataataccagagacagcttctgactgcgaactgtgct 356
Db 181 TATCTGACGACCTGTATATATACCAGAGAACAGCTTTGCTACTGCAATCTGGCT 240
QY 357 tcactgcacccccctccatgacaaatgagtcagagtaagcgcctagatcccttaata 416
Db 241 TCACGTCCATCCCCCTCCATGACAAATGAGTCAGAGTAAGCGCTAGGATCCCTTAATA 300
QY 417 agtttaagcccaagcccgcttcctctctctagcaactgacgttgcagcagcaggtttgg 476
Db 301 AGTTTAAGCCCAAGCCCGCTTCCTCTCTAGCAACTGACGTGCGCAGCCAGGTTTGG 360
QY 477 acatacctcatgtagatgtgtgtgcaactttgacattcctaccattcccaagattaca 536
Db 361 ACATACCTCATGTAGATGTGGTTGTCAACTTTGACATTCCTACCATTCACCAAGGATTACA 420

```

```
QY 537 tccatcgatgtagacagctagagctggcgctcccgaaagcgtattactttgtca 596
|||||
Db 421 TCCATCGAGTAGTCCAGACAGTAGAGCTGGCGCTCCGGAAAGGCTATTACTTTGTCA 480

QY 597 cacagtatgatggaactcttccacgcatgaacacacttaattgggaagaactaccag 656
|||||
Db 481 CACAGTATGATGGAACTCTTCCAGCGCATAGAACACTTAATTGGGAAGAACTACCAG 540

QY 657 gttttcccaacagatgatgaggttatgatctgacagaacgcgtccc 705
|||||
Db 541 GTTTCCTCAACAGGATGATGAGGTATGATGCTGACAGACGCGTGC 589

RESULT 11
LOCUS AI763426 762 bp mRNA EST 20-DEC-1999
DEFINITION wh92c02.x1 NCI_CGAP CLL1 Homo sapiens cDNA clone IMAGE:2388194 3'
similar to TR:009043 009043 KIDNEY-DERIVED ASPARTIC PROTEASE-LIKE
PROTEIN ; mRNA sequence.
ACCESSION AI763426
VERSION AI763426.1 GI:5179093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 762)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1254 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 476.
FEATURES
Location/Qualifiers
1..762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2388194"
/clone_lib="NCI_CGAP CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCATGCTTTTGTGTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 164 a 230 c 231 g 134 t 3 others
ORIGIN

Query Match 27.8%; Score 572.4; DB 24; Length 762;
Best Local Similarity 88.2%; Pred. No. 1.5e-147;
Matches 678; Conservative 0; Mismatches 84; Indels 7; Gaps 5;

QY 1281 tttccattctgtggaagagtgctggcccgatgtagtactgtgagcagggg 1340
|||||
Db 762 TTTCCATTCTGTGGAAGAGTGNCGCCCGCTGGATGNACT-GTGGAGCAGGG 704
```

```
QY 1341 ctattgataagcctgtcttctctcttttaacctcaacagggacccctgaagagcctgga 1400
|||||
Db 703 CTATGGGATAAGCCTGCTTCCTCTTTTACTTCACANGGACCCCTAAGTGGCTGATGGA 644

QY 1401 ggagagctggtcctgggggctggagcccggaacacacacacccctcaaccttcgt 1460
|||||
Db 643 GAAGAGCTGGT-CTGGGGGGCTCAGACCC-GCACACTACATCCACCCCTCACCCTTCGTG 586

QY 1461 ccagtcacggtccctcctactgagagatccacatggagcgtgtgaagtgggccacagg 1520
|||||
Db 585 CCAGTCCACAGTCCCGCCCTACT-GCAGATCCACATGGAGCGTGTGAAGTGGGCTCACGG 527

QY 1521 ctgaactctgtgccaaggcgtgtgctgcacatcctggatgacggcgctccctcatcaca 1580
|||||
Db 526 CTGACTCTCTGTGCCAGGGCTGTGCTGCATCTCTGGATACAGGCACACCTGTATCGTA 467

QY 1581 ggaccactgagagatccgggcccctgcatgcagccatgggggaatcccccttgcgtgct 1640
|||||
Db 466 GGACCCACTGAGGAGATCCGGGCCCTGCATGCAAGCCATTGGGGGAATCCCTTGTCTGGCT 407

QY 1641 ggggagatcatcatcctgtgctcggaataatcccaaaagctcccccgagctccttcttctt 1700
|||||
Db 406 TGGGAGTACATCATCCGGTGTCTCAGAAATCCCAAGCTCCCTCAGTCTCACCCTCATTT 347

QY 1701 gggggggtctggttttaacctcacggcccatgattacgtcatccagactactcgaaatggc 1760
|||||
Db 346 GGGGGGCTCTGGTTTAATCTCAGCGGCCAGGATTAGTCATCCAGTTGCTCAGGGTGAC 287

QY 1761 gtccgctctgtgttccgggtttccaggccctggatgtccctccgcctgcagggcccttc 1820
|||||
Db 286 GTCCGCTCTGCTGTGTCGGGCTTCGGGCTTGGACATCGTTCGGCTCCAGTACCTGTG 227

QY 1821 tggatcctcgtgtagccttcttggggagcgtatgtggcgtcttccagcggggagacatg 1880
|||||
Db 226 TGGATCCTCGGACACCTTTTCTTTGGGGGCGTATGTACCGTCTTCGACCGCGGGGACATG 167

QY 1881 aagacagcgcccggtggcgctggcgctgcactcgcgagcgagcgacctcgatgg 1940
|||||
Db 166 AAGAGCGGCGCAGAGTGGGACTTGGCGCGGCTCGCCCTCGCGAGCGGACCTTGGGAGG 107

QY 1941 ggagagactgcgagcgagcttccccggggtgagcccccaagtgaagcgcatgcagcgcg 2000
|||||
Db 106 CGGAGACCGCGCAGCGCAGTAGTACCGCGGTGCGGCCAGGTGATGCGATGCGATGCCACCG- 46

QY 2001 gtggtcgcgaggtcctgctacccagtaaaatccactatttccattga 2049
|||||
Db 47 --GGTAGCAGAGCTAGCGCTACTCAGTAAATAATCCAAATATTTCCATTGA 1

RESULT 12
LOCUS BF869544 587 bp mRNA EST 17-JAN-2001
DEFINITION IL3-ET0115-091000-286-E09 ET0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF869544
VERSION BF869544.1 GI:12259674
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 587)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,I.G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
```





```

Qy 346 aaatttggttcaactccctccatgacaaatgagtcagagtaagcgctagg 405
|||||
Db 411 AAATCTGGCTTCACTGCATCCCTCCATGACAAATGAGTCAGATAGCGCTAGG 470
Qy 406 atcccttaagttaaggcccaagcccgcttccattctctagcaactgagctggcag 465
|||||
Db 471 ATCCCTTAATAGTTTAAGGGCAAGCGCGCTTCCATTCTCTAGCAACTGAGCTGGCCAG 530
Qy 466 cagagtttgacacatccctcatgtatgtggtgtcaacttt---gacattcctacca 522
|||||
Db 531 CCGAGTTTGGACATACCTCATGTAGATGTGGTTGTCAACCTTTTGACATTTCTACCCAT 590
Qy 523 ttccaagattacatccctcatgtaggtcgaacagctagagctggcgctccgg---aaa 579
|||||
Db 591 TTCCAAGGATTACATCCATCGAGTAGGTGCGAACACGCTAGAGTGGCGCTCCCGGAAAG 650
Qy 580 ggctattactttgtcacacagtagt-attgtggaactcttcacagcgcatagaaacttaa 638
|||||
Db 651 GCTATTTACTTTTGTACACAGATATGTATGTGGGAACCTTTCCAGGCGATAGAACCTTAA 710
Qy 639 ttgggaagaataccaggttttccacacacagagtagtgaggttatgtatgctgacagaa 697
|||||
Db 711 TGGGGAGAAACTACCAGGTTTTCACCCAGGATTGATGAGGGTTGATGCTGACAGAA 769

RESULT 14
LOCUS BE622333 813 bp mRNA EST 20-OCT-2000
DEFINITION 601441119F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3916013 5',
mRNA sequence.
ACCESSION BE622333
VERSION BE622333.1 GI:98933273
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 813)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTO/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9740 row: e column: 06
High quality sequence stop: 664.
FEATURES
source Location/Qualifiers
1..813
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3916013"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 237 a 182 c 203 g 191 t
ORIGIN

Query Match 26.8%; Score 552.4; DB 138; Length 813;
Best Local Similarity 99.8%; Pred. No. 5.4e-142;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 152 ctgaagaatcctgtgaaatgtgcggttctcttaataaccagacagtgtaaaaaatacag 211
|||||
Db 1 CTGAAGAATCCTGTGAATGTGCCGTTTCTCTAAATACCAGACAGTTGAAAAATTACAG 60
Qy 212 caatattatttttattccctctcaaaattcaaggatacctaccctggtttattctataat 271
|||||
Db 61 CAATATTATATTTTATTCCTCTAAATTTCAAGGATACCTAGCTGGTTTATATCTTAAT 120
Qy 272 gaattggctggaacctcctttatgatattctgcagcacctgtaataataaccagagaaca 331
|||||
Db 121 GAATTGGCTGGAAACCTCTTTATGATATTTCTGCAGCACCTGTAAATATACCCAGAGAAC 180
Qy 332 gctttgctactgcgaatacttggcttcaactgcacatccctccatcgacacaaatgagtcag 391
|||||
Db 181 GCTTTGCTACTCGAAATCTTGGCTTCACTGCATCCCTCCATGCACAAATGAGTCAG 240
Qy 392 agtaagcgcctaggatcccttaataagtttaaggccaagggcccgcttccattctctaga 451
|||||
Db 241 AGTAAGCGCTAGGATCCCTTAATAGTTTAAGGGCAAGCGCGCTTCCATTCTCTAGCA 300
Qy 452 actgagcttgcagcgcgaggtttggacatacctcatgtatgtagtgggttgcactttgac 511
|||||
Db 301 ACTGAGCTTGCACGCGAGGTTTGGACATACCTCATCTAGATGTGGTTGCACTTTGAC 360
Qy 512 attctaccattcccaaggattacatccatcgtagtgcgaacagctagagctgggcgc 571
|||||
Db 361 ATTCTACCCATTCCAAGGATTACATCGAGTAGGTGGAACAGCTAGAGTGGCGC 420
Qy 572 tcggaaaggctattacttttgcacacagtagtgcgaactcttcacagcgcatagaa 631
|||||
Db 421 TCCGGAAGGCTATTACTTTTGTACACAGATATGATGTGGAACCTCTTCCAGCGCATAGAA 480
Qy 632 cacttaattgggaagaacctaccaggttttccacacagagtagtgaggttatgctgctg 691
|||||
Db 481 CACTTAATTGGGAAGAACTACACAGGTTTCCACACAGAGTAGTAGGTTATGATGCTG 540
Qy 692 acagaacgcgtccc 705
|||||
Db 541 ACAGAAGCGGTCGC 554

RESULT 15
LOCUS AA744560/c
DEFINITION ny79dl1.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1384501 3'
similar to TR:009043 009043 KDAP-1.; mRNA sequence.
ACCESSION AA744560
VERSION AA744560.1 GI:2783324
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 837)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 942 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 450.
Location/Qualifiers
FEATURES

```

Db	173	GAAGAGCGGCACGAGTGGGACTGGCGCGCTCGCCCTCCGGAGCGGACCTGGGAGC	114
Qy	1940	ggagagactgcgacgcagtcagttcccgggtgcagcccaagtgaa'gcgcatgcgcagcg	1999
Db	113	GCAGACGACGAGCAGGCGCGAG-TACCGCGGTGCGCGCAGGTTGATGCGCACCG	58
Qy	2000	ggtggtgcgaggtcctgctaccacgataaaatccactattccattgaaaaa	2059
Db	57	GGTAGCAGTAG- ---CGCTACTCAGTAAAAATCCAAATATTTCCATTGAAAAA	2
Qy	2060	a	2060
Db	1	A	1

Search completed: November 17, 2001, 13:34:21  
Job time: 3420 sec

Query Match	26.7%;	Score 550;	DB 11;	Length 837;
Best Local Similarity	84.7%;	Pred. No. 2.5e-141;		
Matches	712;	Conservative	0;	Mismatches 116;
				Indels
				13;
				Gaps
				8;
QY	1222	tttggagaccagcctggtcttcgtcttggcattttggcattttgatggatatatgg-gcctcggt	1280	
DB	830	TTTGAAACCAACCTTCTCTCACTGTTTCCGCGCCGCTGGGATATTGGCGCCTCGTT	771	
QY	1281	tcccaattctgtctgtgaaagaggttcgccccccgatgatgtactggtggaacaggg	1340	
DB	770	TTCCCAATTCCTGTGTGGAGAGTTTGGCCCCCGTGGATTACTTGGTGGAGACAGGC	711	
QY	1341	ctattggataagcctgtcttctcttttacctcaacaggaccctgaagcctgatgtgga	1400	
DB	710	CTATTGGATAAGCTGTCTTCTCCCTTTACTTTCAACAGGCACCTGAAGTGGCTGAT-GA	652	
QY	1401	gagagctggtccctgggggctcggaacccggcacactacatccca-cccccaacctcgt	1459	
DB	651	GACAGCTGGTCTTGGGGGCTTCAGACCCGGACACTACATCCACCCCTCACCTTCGT	592	
QY	1460	gcagtcacaggtccctgcctacttggcagatccacatggagcgtgaaagtggcccagg	1519	
DB	591	GCAGTCACAGTCCCGCCCTACTTGGCAGATCCACATGGAGCGTGTGAAGTGGGCTCAG	532	
QY	1520	gctgaactctctgtgccaaaggctgtctgccaactctggatacagggcacgtccctcatcac	1579	
DB	531	GCTGACTCTCTGTGCCACAGGCTGTGTGTCATCTCTGGATACAGGCACACCTGTCTACGT	472	
QY	1580	agaccacactgagagatccggccctgatcgagcattgggggaatcccccttgcctggc	1639	
DB	471	AGACCCACTTGAAGAGATCCGGGCCCTGTGATGACGCCATT-GGGGAATCCCTTGTGTGC	413	
QY	1640	tggggagtaacatcatcctgtgctcggaatccccaaagctccccgagctcctctctct	1699	
DB	412	TGSGGAGTACATCATCCGGTGTCTCAGAAATCCAAAGCTCCCCGCAGTCTCACTCAT	353	
QY	1700	tgggggggtctggtttaacctcaagcccactgattactgcattccagactactcgaatgg	1759	
DB	352	TGSGGGGGTCTGGTTTAAATCTCAGGCCCAGGATTACGTATCCATCCAGTTTGGCTCAGGGTCA	293	
QY	1760	cgtccgcctctctgtctcggtttccagccctgatgtccctccgcctgcagggccctt	1819	
DB	292	CGTCCGCCCTCTGCTGTCCGGCTTCGGGGCCCTTGGG-CAATCTTCGCCTCCAGTACCTGT	234	
QY	1820	cttggatcctcgttgaactcttcttgggaacgttatgtgccgtcttcgcaccgcgggacat	1879	
DB	233	GTFAATCTCTCGGCACGTTTCTTGGGGCGGTATGTGACCGTCTTCGACCCGGGGACAT	174	
QY	1880	gaagagacagcccggttggcctgcgcgcgcctgcactgcggagcgcgacctcggtat	1939	

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:54:41 ; Search time 4216.79 Seconds  
(without alignments)  
1991.797 Million cell updates/sec

Title: US-09-700-770-6  
Perfect score: 543  
Sequence: 1 ccggcgtgaggggaggg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: gb\_bal:\*
  - 2: gb\_ba2:\*
  - 3: gb\_ba3:\*
  - 4: gb\_in1:\*
  - 5: gb\_in2:\*
  - 6: gb\_in3:\*
  - 7: gb\_om:\*
  - 8: gb\_ov:\*
  - 9: gb\_pat1:\*
  - 10: gb\_pat2:\*
  - 11: gb\_ph:\*
  - 12: gb\_pl1:\*
  - 13: gb\_pl2:\*
  - 14: gb\_pl3:\*
  - 15: gb\_pl4:\*
  - 16: em\_bal:\*
  - 17: em\_ba2:\*
  - 18: em\_fun:\*
  - 19: em\_htgo\_hum:\*
  - 20: em\_htgo\_inv:\*
  - 21: em\_htgo\_rod:\*
  - 22: em\_htg\_hum1:\*
  - 23: em\_htg\_hum2:\*
  - 24: em\_htg\_hum3:\*
  - 25: em\_htg\_hum4:\*
  - 26: em\_htg\_hum5:\*
  - 27: em\_htg\_hum6:\*
  - 28: em\_htg\_hum7:\*
  - 29: em\_htg\_hum8:\*
  - 30: em\_htg\_inv1:\*
  - 31: em\_htg\_inv2:\*
  - 32: em\_htg\_other:\*
  - 33: em\_htg\_rod:\*
  - 34: em\_hum1:\*
  - 35: em\_hum2:\*
  - 36: em\_hum3:\*
  - 37: em\_hum4:\*
  - 38: em\_hum5:\*
  - 39: em\_hum6:\*
  - 40: em\_hum7:\*
  - 41: em\_in:\*
  - 42: em\_om:\*
  - 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_vl:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_vl1:\*
- 59: gb\_vl2:\*
- 60: gb\_vtg1:\*
- 61: gb\_vtg2:\*
- 62: gb\_vtg3:\*
- 63: gb\_vtg4:\*
- 64: gb\_vtg5:\*
- 65: gb\_vtg6:\*
- 66: gb\_vtg7:\*
- 67: gb\_vtg8:\*
- 68: gb\_vtg9:\*
- 69: gb\_vtg10:\*
- 70: gb\_vtg11:\*
- 71: gb\_vtg12:\*
- 72: gb\_vtg13:\*
- 73: gb\_vtg14:\*
- 74: gb\_vtg15:\*
- 75: gb\_vtg16:\*
- 76: gb\_vtg17:\*
- 77: gb\_vtg18:\*
- 78: gb\_vtg19:\*
- 79: gb\_vtg20:\*
- 80: gb\_vtg21:\*
- 81: gb\_vtg22:\*
- 82: gb\_vtg23:\*
- 83: gb\_vtg24:\*
- 84: gb\_vtg25:\*
- 85: gb\_pr1:\*
- 86: gb\_pr2:\*
- 87: gb\_pr3:\*
- 88: gb\_pr4:\*
- 89: gb\_pr5:\*
- 90: gb\_pr6:\*
- 91: gb\_pr7:\*
- 92: gb\_pr8:\*
- 93: gb\_pr9:\*
- 94: gb\_ro1:\*
- 95: gb\_ro2:\*
- 96: gb\_in4:\*
- 97: gb\_pr10:\*
- 98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match %	Length DB ID		
1	471	86.7	471 97	HUM2B52D10	AF086152 Homo sapi
C 2	241.4	44.5	96482 67	AC022095	AC022095 Homo sapi
C 3	241.4	44.5	168347 69	AC025336	AC025336 Homo sapi
C 4	60.4	11.1	40104 3	SCF11	AL132662 Streptomy
C 5	60.2	11.1	34611 3	SCD63A	AL356832 Streptomy
C 6	56	10.3	208936 61	AC010821	AC010821 Homo sapi
C 7	55.8	10.3	166539 62	AC011402	AC011402 Homo sapi
C 8	55.4	10.2	33001 3	SCC123	AL136518 Streptomy

c 9 52.8 9.7 26423 3 SC8612  
 c 10 52.4 9.7 52884 73 AC068018  
 c 11 52.2 9.6 149719 78 AC090977  
 c 12 52 9.6 12924 1 AE002059  
 c 13 51.2 9.4 160949 62 AC011352  
 c 14 51.2 9.4 207747 62 AC011334  
 c 15 51 9.4 37931 3 SC010  
 c 16 50.8 9.4 3957 9 A45258  
 c 17 50.8 9.4 154746 59 H5V2HG52  
 c 18 50.8 9.4 154746 59 H5V2HG52  
 c 19 50.4 9.3 37898 3 SCE34  
 c 20 50 9.2 65225 75 AC079109  
 c 21 50 9.2 77457 2 AF210249  
 c 22 50 9.2 187738 71 AC027682  
 c 23 49.8 9.2 1875 89 AF169017  
 c 24 49.8 9.2 1900 89 AF289021  
 c 25 49.8 9.2 1930 89 AF289024  
 c 26 49.8 9.2 1948 89 AF289022  
 c 27 49.8 9.2 4523 59 MMULGENES  
 c 28 49.6 9.1 146081 13 AP001366  
 c 29 49.4 9.1 20535 3 SC2G4  
 c 30 49.2 9.1 154672 83 AP003512  
 c 31 48.6 9.0 16074 68 AC023682  
 c 32 48.4 8.9 34023 3 SC2G2  
 c 33 48.4 8.9 299003 75 AC078780  
 c 34 48.2 8.9 34182 3 SC111  
 c 35 48.2 8.9 198873 67 AC022811  
 c 36 48 8.8 40790 3 MTCY493  
 c 37 47.8 8.8 2586 3 THTTRPBA  
 c 38 47.8 8.8 7713 7 RABBOCBIII  
 c 39 47.8 8.8 249687 63 AC015693  
 c 40 47.6 8.8 1831 89 AF289023  
 c 41 47.6 8.8 126427 77 AC087096  
 c 42 47.6 8.8 160536 71 AC040953  
 c 43 47.6 8.8 167540 64 AC016084  
 c 44 47.4 8.7 10610 1 AE002029  
 c 45 47.4 8.7 10910 1 AE004480

## ALIGNMENTS

RESULT 1  
 HUMZB52D10 471 bp mRNA PRI 29-AUG-1998  
 LOCUS Homo sapiens full length insert cDNA clone ZB52D10.  
 DEFINITION AF086152  
 ACCESSION AF086152.1 GI:3483497  
 VERSION FLI\_CDNA.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 471)  
 Woessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,  
 Martel,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,  
 Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,  
 Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,  
 Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behymer,K., Hillier,L.,  
 Willson,R. and Waterston,R.  
 Full Clone Sequencing of the Longest Available Member from Each  
 Unigene Cluster  
 Unpublished  
 2 (bases 1 to 471)  
 Waterston,R.  
 Direct Submission  
 Submitted (24-AUG-1998) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 SUBMITTED BY:  
 Genome Sequencing Center  
 Department of Genetics  
 Washington University  
 St. Louis MO 63108, USA

http://genome.wustl.edu/gsc  
 mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

FEATURES  
 source  
 1. .471  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Soares.fetal\_lung\_NbHL19"  
 /clone="IMAGE:307219" 73 t

BASE COUNT 78 a 171 c 149 g 73 t  
 ORIGIN

Query Match 86.7%; Score 471; DB 97; Length 471;  
 Best Local Similarity 100.0%; Pred. No. 8e-69;  
 Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 cgcgccccgagccccgcgcacgaagctgcgcgcctcctcctgggctctcgtgcgcctg 131  
 Db 1 CGCGCCCGAGCCCGCCCGGCATGAAGCTCGCGCCCTCTCTGGGCTCTGGTGGCCCTG 60  
 QY 132 tcttcgagctccgctgctgttctttagtggttcggccaaagcctgtggccagcctgtc 191  
 Db 61 TCCTGACGCTCCGCTGCTGCTTCTTTAGTGGGCTCGGCCAAGCCTGTGGCCCAACCTGTC 120

QY 192 gctgcgctggagtcggcgcgagcgccggcgccgacctggccacccctcggcacc 251  
 Db 121 GCTGCGCTGGAGTCGGCGCGGAGCGCGGCGGACCTGGCCACCCCTCGGCACC 180

QY 252 ctaaacccgctgaagctcctgctgagcagcctgggcatccccctgaaccacctcatag 311  
 Db 181 CTCAACCCGCTGAAGCTCTCTGCTGAGCAGCCTGGGCATCCCGTGAACCACTCATAG 240

QY 312 ggctcccgagaagtgtgtgctgagctgggtggtcccgagcgtgggcccgtgaagccctg 371  
 Db 241 GGCTCCAGAGTGTGTGGCTGAGCTGGGTCCCGAGCCCGGGGGCGGGAAGGCCCTG 300

QY 372 aaggccctgctggggccctgacagtgtttggctgagcgcgagactggagcatctacac 431  
 Db 301 AAGGCCCTGCTGGGGCCCTGACAGTGTGCTGAGCGGAGACTGGAGCATCTACACCT 360

QY 432 gagacaacgctgcccccccgagggtgaaacccccgcgcgcgggagagacctcca 491  
 Db 361 GAGACAAGACGCTGCCACCCCGAGAGGCTGAAACCCCGCGCGGGAGGACCGTCCA 420

QY 492 tcccttcccccgccctctcctaataacgtgttaagagcaaaaaaaa 542

Db 421 TCCCTTCCCGGCCCTCTCAATAACGTGTTAAGAGCAAAAAAAA 471

RESULT 2  
 AC022095/c

LOCUS

DEFINITION

AC022095

AC022095

AC022095.4

HTG: HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

ORGANISM

AC022095 96482 bp DNA HTG 06-MAY-2000  
 Homo sapiens chromosome 5 clone CTF-36B8, WORKING DRAFT SEQUENCE,  
 14 unordered pieces.

AC022095  
 AC022095.4 GI:7711676  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 human.  
 Homo sapiens



Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L6686

Center clone name: 451\_H23

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 150422 bases at least Q40

Consensus quality: 159524 bases at least Q30

Consensus quality: 163013 bases at least Q20

Insert size: 165247; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 1389: contig of 1389 bp in length
1390 1489: gap of 100 bp
1490 3130: contig of 1641 bp in length
3131 3230: gap of 100 bp
3231 4942: contig of 1712 bp in length
4943 5042: gap of 100 bp
5043 6981: contig of 1939 bp in length
6982 7081: gap of 100 bp
7082 8708: contig of 1627 bp in length
8709 8808: gap of 100 bp
8809 10286: contig of 1478 bp in length
10287 10386: gap of 100 bp
10387 12212: contig of 1826 bp in length
12213 12312: gap of 100 bp
12313 14658: contig of 2346 bp in length
14659 14758: gap of 100 bp
14759 17941: contig of 3183 bp in length
17942 18041: gap of 100 bp
18042 21297: contig of 3256 bp in length
21298 21397: gap of 100 bp
21398 24992: contig of 3595 bp in length
24993 25092: gap of 100 bp
25093 27768: contig of 2676 bp in length
27769 27868: gap of 100 bp
27869 31188: contig of 3320 bp in length
31189 31288: gap of 100 bp
31289 33714: contig of 2426 bp in length
33715 33814: gap of 100 bp
33815 37277: contig of 3463 bp in length
37278 37377: gap of 100 bp
37378 42302: contig of 4925 bp in length
42303 42402: gap of 100 bp
42403 47816: contig of 5414 bp in length
47817 47916: gap of 100 bp
47917 52586: contig of 4670 bp in length
52587 52686: gap of 100 bp
52687 56567: contig of 3881 bp in length
56568 56667: gap of 100 bp
56668 61557: contig of 4890 bp in length
61558 61657: gap of 100 bp
61658 66724: contig of 5067 bp in length
66725 66824: gap of 100 bp
66825 71568: contig of 4744 bp in length
71569 71668: gap of 100 bp
71669 76578: contig of 4910 bp in length
76579 76678: gap of 100 bp
76679 83312: contig of 6634 bp in length
83313 83412: gap of 100 bp

```

```

83413 90053: contig of 6641 bp in length
90054 90153: gap of 100 bp
90154 99426: contig of 9273 bp in length
99427 99526: gap of 100 bp
99527 108015: contig of 8489 bp in length
108016 108115: gap of 100 bp
108116 118144: contig of 10029 bp in length
118145 118244: gap of 100 bp
118245 130466: contig of 12224 bp in length
130469 130568: gap of 100 bp
130569 142239: contig of 11671 bp in length
142240 142339: gap of 100 bp
142340 157135: contig of 14796 bp in length
157136 157235: gap of 100 bp
157236 168347: contig of 11112 bp in length.

```

#### FEATURES

```

Location/Qualifiers
source
1..168347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"
/clone="RP11-451H23"
/clone_lib="RPC1-11 Human Male BAC"
1..1389
/note="assembly_fragment"
1490..3130
/note="assembly_fragment"
3231..4942
/note="assembly_fragment"
5043..6981
/note="assembly_fragment"
7082..8708
/note="assembly_fragment"
8809..10286
/note="assembly_fragment"
10387..12212
/note="assembly_fragment"
12313..14658
/note="assembly_fragment"
14759..17941
/note="assembly_fragment"
18042..21297
/note="assembly_fragment"
21398..24992
/note="assembly_fragment"
25093..27768
/note="assembly_fragment"
27869..31188
/note="assembly_fragment"
31289..33714
/note="assembly_fragment"
33815..37277
/note="assembly_fragment"
37378..42302
/note="assembly_fragment"
42403..47816
/note="assembly_fragment"
47917..52586
/note="assembly_fragment"
52687..56567
/note="assembly_fragment"
56668..61557
/note="assembly_fragment"
61658..66724
/note="assembly_fragment"
66825..71568
/note="assembly_fragment"
71669..76578
/note="assembly_fragment"
76679..83312
/note="assembly_fragment"
83413..90053
/note="assembly_fragment"

```

#### misc\_feature

```

misc_feature
1490..3130
/note="assembly_fragment"
3231..4942
/note="assembly_fragment"
5043..6981
/note="assembly_fragment"
7082..8708
/note="assembly_fragment"
8809..10286
/note="assembly_fragment"
10387..12212
/note="assembly_fragment"
12313..14658
/note="assembly_fragment"
14759..17941
/note="assembly_fragment"
18042..21297
/note="assembly_fragment"
21398..24992
/note="assembly_fragment"
25093..27768
/note="assembly_fragment"
27869..31188
/note="assembly_fragment"
31289..33714
/note="assembly_fragment"
33815..37277
/note="assembly_fragment"
37378..42302
/note="assembly_fragment"
42403..47816
/note="assembly_fragment"
47917..52586
/note="assembly_fragment"
52687..56567
/note="assembly_fragment"
56668..61557
/note="assembly_fragment"
61658..66724
/note="assembly_fragment"
66825..71568
/note="assembly_fragment"
71669..76578
/note="assembly_fragment"
76679..83312
/note="assembly_fragment"
83413..90053
/note="assembly_fragment"

```

```

misc_feature      90154..99426
                  /note="assembly_fragment"
misc_feature      99527..108015
                  /note="assembly_fragment"
misc_feature      108116..118144
                  /note="assembly_fragment"
misc_feature      118245..130468
                  /note="assembly_fragment"
misc_feature      130569..142239
                  /note="assembly_fragment"

Query Match      44.5%; Score 241.4; DB 69; Length 168347;
Best Local Similarity 99.6%; Pred. No. 5.3e-32;
Matches 242; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 142 ccgctgctgttttaagtgcctgcccgaagcctgtgcccagcctctgctgcgtgg 201
      |||
Db 144047 CAGCTGCTGCTTTCTTATGTGGCTGCGGCAAGCCCTGTGCCAGCCTGTGCTGCGCTGG 143988

Qy 202 agtcggcgagcggcgccggcggaacccctggccaaacccctcgccaccctcaaccgc 261
      |||
Db 143987 AGTCGGCGGAGCGCGCGCGGACCCCTGGCCAAACCCCTCGGCACCCCTCAACCCGC 143928

Qy 262 tgaagctcctgctgagcagcctgggcctcccgctggaacacacctatagaggctcccgaga 321
      |||
Db 143927 TGAAGCTCCTGCTGAGCAGCCTGGGCATCCCGCTGAACCACTCATAGAGGGCTCCCGAGA 143868

Qy 322 agtgtgtgctgagctggtcccccagcgcgtggggccgtgaagccctgaagccctgc 381
      |||
Db 143867 AGTGTGTGCTGAGCTGGTGTCCCGAGCCGCGGGGCGCGTGAAGGCCCTGAGGCCCTGC 143808

Qy 382 tgg 384
      |||
Db 143807 TGG 143805

RESULT 4
LOCUS      SCF11 40104 bp DNA BCT 26-OCT-1999
DEFINITION Streptomyces coelicolor cosmid F11.
ACCESSION AL132662
VERSION AL132662.1 GI:6137022
KEYWORDS 3-oxoacyl-(acyl-carrier-protein) synthase; acyl carrier protein;
acyltransferase; alpha-galactosidase; beta-hexosaminidase; cold
shock protein; gluconolactonase precursor; histidine kinase
protein; hydrolase; oxidoreductase; peptidase; pyruvate
carboxylase; regulatory protein; response regulator; sugar
hydrolase; sugar transporter; transcriptional regulator;
transmembrane protein.
SOURCE      Streptomyces coelicolor A3(2).
ORGANISM    Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Streptomyces.
REFERENCE   1 (bases 1 to 40104)
AUTHORS     Redenbach,M., Kieser,H.M., Denapate,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE       A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
MOL. MICROBIOL. 21 (1), 77-96 (1996)
JOURNAL     97000351
MEDLINE     2 (bases 1 to 40104)
REFERENCE   Seeger,K.J. and Harris,D.
JOURNAL     Unpublished
AUTHORS     3 (bases 1 to 40104)
Corden,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE       Direct Submission
JOURNAL     Submitted (26-OCT-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded

```

by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid F11 overlaps with cosmid F6 on the AseI-F genomic restriction fragment.

FEATURES

Location/Qualifiers	Source
1..40104	
/organism="Streptomyces coelicolor A3(2)"	
/strain="A3(2)"	
/db_xref="taxon:100226"	
/clone="cosmid F11"	
1..100	gene
/gene="SCF11.01"	
<1..100	CDS
/gene="SCF11.01"	
/note="SCF11.01, possible tetR family transcriptional regulator, partial CDS, len: >32 aa. previously annotated as Streptomyces coelicolor SCF6.17, possible tetR family transcriptional regulator, len: >193 aa. Similar to several e.g. Streptomyces coelicolor TR:CAB56142 (EMBL; AL117669) putative transcriptional regulator SCF12.15 (188 aa), fasta scores opt:446 z-score: 521.2 E(): 1.2e-21 43.5% identity in 184 aa overlap and Streptomyces glaucescens SW:TCMR_STRGA (EMBL; M80674) tetracenomycin C transcriptional repressor (226 aa), fasta scores opt: 202 z-score: 242.5 E(): 3.8e-06 29.3% identity in 164 aa overlap."	
/codon_start=2	
/transl_table=11	
/product="putative tetR family transcriptional regulator (partial)"	
/protein_id="CAB59578.1"	
/db_xref="GI:6137023"	
/translation="IDGGRERPLRLAECDFVLTLSARGAASAG"	
1..100	misc_feature
/gene="SCF11.01"	
/note="nominal overlap with Streptomyces coelicolor cosmid ST6"	
254..257	RBS
276..1067	gene
/gene="SCF11.02"	
276..1067	CDS
/gene="SCF11.02"	
/note="SCF11.02, probable oxidoreductase, len: 262 aa; similar to TR:O54197 (EMBL:AJ000671) Streptomyces clavuligerus clavulanate-9-aldehyde reductase, 247 aa; fasta scores: opt: 501 z-score: 567.7 E(): 3e-24; 39.1%	

	identity in 238 aa overlap. Contains Pfam match to entry PF00106 adh_short, short chain dehydrogenase and prosite match to entry PS0061 Short-chain dehydrogenases/reductases family signature. also similar to Streptomyces coelicolor SCF6.15c; 245 aa; fasta scores: opt: 818 z-score: 805.2 E():0; 56.5% identity in 237 aa overlap" /codon_start=1 /transl_table=11 /product="putative oxidoreductase" /protein_id="CAB59579.1" /db_xref="GI:6137024" /translation="MSRFDQAQWRTTHRKDVIMTQTSKVVLVTGASSGIGEATLRLLAADGHRFLGARRTERLEKLAARIAEDGGTAGVRLRDLDAADVRAFSVAARWGLALDVI VNNAGQSLSPLEKLAERMLDVNVRVHLGIIAAALPVRMAQGGGHVVNLGAYEVPTAAVYCATFAVRALSEGRLQESAGDIRVSYSPGVSTSELADSIDSPRAAREDKMTTYSVAVPAASATAATAFAVSRPAEVDVNEIVVRPAASAQ" 348. .899 /gene="SCF11.02" /note="Pfam match to entry PF00106 adh_short, short chain dehydrogenase, score 248.20, E-value 1.2e-70" 753. .839 /gene="SCF11.02" /note="PS0061 Short-chain dehydrogenases/reductases family signature" complement(1095. .1532) /gene="SCF11.03c" complement(11164. .1460) /gene="SCF11.03c" complement(11164. .1460) /note="SCF11.03c, probable transcriptional regulator, len: 145 aa; similar to SW:ASNC_ECOLI (EMBL:AE000451) Escherichia coli, regulatory protein AsnC. 152 aa; fasta scores: opt: 222 z-score: 287.1 E(): 1.3e-08; 26.9% identity in 145 aa overlap. Contains Pfam match to entry PF01037 ASNC_trans_reg, AsnC family" /codon_start=1 /transl_table=11 /product="putative transcriptional regulator" /protein_id="CAB59580.1" /db_xref="GI:6137025" /translation="MDEIDRRLLALQQDAPRYAALGGEVGLSGAHAHRYVRKLRREQKVRITTVDPBALDRGVLAFLVVDSTAMWGDRAEDFAAPFIOFAHVIAGSAALVKTATRELQVLDRLRYLAIYDVGSTQATVLTETFERPLDAG" complement(1164. .1460) /gene="SCF11.03c" /note="Pfam match to entry PF01037 ASNC_trans_reg, AsnC family, score 52.20, E-value 1.1e-11" 1608. .1612 1617. .2525 1617. .2525 /gene="SCF11.04" /gene="SCF11.04" /note="SCF11.04, possible gluconolactonase, len: 302 aa; similar to SW:GNL_ZYMMO (EMBL:X67189) Zymomonas mobilis gluconolactonase precursor (EC 3.1.1.17) D-glucono-delta-lactone lactonohydrolase; fasta scores: opt: 326 z-score: 392.6 E(): 1.7e-14; 28.7% identity in 258 aa overlap" /codon_start=1 /transl_table=11 /product="possible gluconolactonase precursor" /protein_id="CAB59581.1" /db_xref="GI:6137026" /translation="MSTDGPYILDDRFRTGCRANGNRLEVLVDGCRWAGPLVLPALWQLVMSDIPNDRILRWRDEATSGVGRAPAGHSNGNTLDRQGLVTCCEQNRVFT EPDGRVTYLAERDGRRLSPNDVSRSDPTGWTFSDPFGITVSEGHRASEIGACN VYRIDPVSQGVRLADGEDGPNGLITPDERRLRFVSDSRAAKIHAFIDREDGTLSDGK VRAEGRGVTFDNIIFDEDEGRULWAGHDGVCYDPTGTLIGLRUPPEVSVNAFVGG KNNRIFTATTSILYSUWVSZGAPRL" 2656. .3471 /gene="SCF11.05" 2656. .3471 /gene="SCF11.05"
misc_feature	
misc_feature	
gene	
CDS	
misc_feature	
RBS	
gene	
CDS	
gene	
CDS	



REFERENCE	2 (bases 1 to 34611)	
AUTHORS	Brown,S.P. and Harris,D.	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 34611)	
AUTHORS	Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-MAY-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge	
COMMENT	DB10 JSA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK	
COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: <a href="http://www.sanger.ac.uk/Projects/S_coelicolor/">http://www.sanger.ac.uk/Projects/S_coelicolor/</a> ) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Blobb et al., Gene 30:157-66(1984) as implemented at <a href="http://www.nh.hg.jp/jp/">http://www.nh.hg.jp/</a> jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. - IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid D63A . Location/Qualifiers 1..34611 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid D63A" complement(1..106) /gene="SCD63A.01c" misc_feature 1..100 /note="nominal overlap qwith Streptomyces coelicolor cosmid SCD63" complement(<1..106) /gene="SCD63A.01c" /note="SCD63A.01c, possible integral membrane protein (fragment), len: >35 aa" /codon_start=1 /transl_table=11 /product="putative integral membrane protein" /protein_id="CAB92653.1" /db_xref="GI:8218191" /translation="MTDHEAATGPGSGPRPPGAGPQDAAPAADAG" 297..1589 /gene="SCD63A.02" 297..1589 /gene="SCD63A.02" /note="SCD63A.02, possible two-component system sensor kinase, len: 430 aa; C-terminal domain similar to TR:CAB51969 (EMBL:AL109661) Streptomyces coelicolor putative two-component system sensor kinase SC6E10.15c, 413 aa; fasta scores: opt: 168 z-score: 198.3 E(): 0.0014;	
FEATURES	Source	
gene		
misc_feature		
CDS		
gene		
CDS		





```

QY 404 ctgagccgagactggagcattctacacctgagagacagacgctgcccacccgcgagggctg 463
Db 100275 CAGGGGGGGCGGNNNNNNCCCCCGCGCGGCGGGGGCGCGCGGNNNGNNGCGGN 100216
QY 464 aaaaacccgcgagggagagcgtctcccttcccttcccccgcctctca 515
Db 100215 NCCCCNCCGNNNCCGCCNCCNCGGCCCGCCCCCGGGCCCGGGGGGCCCCCA 100164

RESULT 7
AC011402
LOCUS Homo sapiens chromosome 5 clone CTB-38B5, WORKING DRAFT SEQUENCE, 3
DEFINITION ordered pieces.
AC011402
VERSION AC011402.7 GI:10305126
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166539)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166539)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 26, 2000 this sequence version replaced gi:9256288.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 78825, H263
Center clone name: CIT978SKB_38B5
-----
Summary Statistics
Consensus quality: 165360 bases at least Q40
Consensus quality: 166150 bases at least Q30
Consensus quality: 166324 bases at least Q20
Estimated insert size: 140000; pulse field gel estimation
Estimated insert size: 168489; sum-of-ctigs estimation
Quality coverage: 8.68 in Q20 bases; pulse field gel estimation
Quality coverage: 7.3 in Q20 bases; sum-of-ctigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 102987: contig of 102987 bp in length
* 102988 103087: gap of unknown length
* 103088 147077: contig of 43990 bp in length
* 147078 147177: gap of unknown length
* 147178 166539: contig of 19362 bp in length.
Location/Qualifiers
1. 166539
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-38B5"
/clone_lib="CalTech human BAC library B"
50180 a 31184 c 31419 g 51645 t 2111 others

```

```

Query Match 10.3%; Score 55.8; DB 62; Length 166539;
Best Local Similarity 61.2%; Pred. No. 0.27;
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 229 ccctggcccaacccccctcgccaccctcaaccgcgtgaagctcctcgtgagcagcctgggca 288
Db 113702 CTCTGGACAACTTCCTCCCTTTATGGATCATTAAGAGCTTCTTCTGAAAACCTCTGGCA 113761

QY 289 tcccctgaaccacctcatagaggctcccagagctgtgctgagctggtgtcccccagg 348
Db 113762 TTTCTGTTGACACCTTCTGGAGGGGCTAAGAGAGTGTTAAATGAGCTGGACACGAGG 113821

QY 349 ccgtggggcgctgaagccctgaagg 375
Db 113822 CTTCTGAAGCTGTGAAGAAACTGCTGG 113848

RESULT 8
SCC123/c
LOCUS Streptomyces coelicolor cosmid C123.
DEFINITION Streptomyces coelicolor cosmid C123.
ACCESSION AL136518
VERSION AL136518.1 GI:6714668
KEYWORDS 30S ribosomal protein S20; DNA-binding protein; gamma-glutamyl
phosphate reductase; glutamate 5-kinase; glycosyl transferase;
integral membrane protein; leucyl-tRNA synthetase; leus; membrane
protein; oxidoreductase; phosphoglycerate mutase; proA; proB; rspT;
secreted protein; tRNA.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 33001)
AUTHORS Redenbach,M., Kieser,H.M., Denapaitte,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
REFERENCE 2 (bases 1 to 33001)
AUTHORS Brown,S.P. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 33001)
AUTHORS Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S.coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.gov/jp/
Jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the

```







Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Vieler, J., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## TITLE

## JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
-----  
Project Information  
Center project name: LI0182  
Center clone name: 46\_L\_10  
-----

\* NOTE: This record contains 66 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
708 807: contig of 707 bp in length  
808 1505: contig of 698 bp in length  
1506 1605: gap of 100 bp  
1606 2303: contig of 698 bp in length  
2304 2403: gap of 100 bp  
2404 3111: contig of 708 bp in length  
3112 3211: gap of 100 bp  
3212 3930: contig of 719 bp in length  
3931 4030: gap of 100 bp  
4031 4731: contig of 701 bp in length  
4732 4831: gap of 100 bp  
4832 5543: contig of 712 bp in length  
5544 5643: gap of 100 bp  
5644 6337: contig of 694 bp in length  
6338 6437: gap of 100 bp  
6438 7148: contig of 711 bp in length  
7149 7248: gap of 100 bp  
7249 7946: contig of 698 bp in length  
7947 8046: gap of 100 bp  
8047 8738: contig of 692 bp in length  
8739 8838: gap of 100 bp  
8839 9524: contig of 686 bp in length  
9525 9624: gap of 100 bp  
9625 10323: contig of 699 bp in length  
10324 10423: gap of 100 bp  
10424 11110: contig of 687 bp in length  
11111 11210: gap of 100 bp  
11112 11910: contig of 700 bp in length  
11911 12010: gap of 100 bp  
12011 12704: contig of 694 bp in length  
12705 12804: gap of 100 bp  
12805 13492: contig of 688 bp in length  
13493 13592: gap of 100 bp  
13593 14282: contig of 690 bp in length  
14283 14382: gap of 100 bp  
14383 15078: contig of 696 bp in length  
15079 15178: gap of 100 bp  
15179 15881: contig of 703 bp in length

15882 15981: gap of 100 bp  
15982 16683: contig of 702 bp in length  
16684 16783: gap of 100 bp  
16784 17504: contig of 721 bp in length  
17505 17604: gap of 100 bp  
17605 18298: contig of 694 bp in length  
18299 18398: gap of 100 bp  
18399 19087: contig of 689 bp in length  
19088 19187: gap of 100 bp  
19188 19871: contig of 684 bp in length  
19872 19971: gap of 100 bp  
19972 20661: contig of 690 bp in length  
20662 20761: gap of 100 bp  
20762 21462: contig of 701 bp in length  
21463 21562: gap of 100 bp  
21563 22265: contig of 703 bp in length  
22266 22365: gap of 100 bp  
22366 23065: contig of 700 bp in length  
23066 23165: gap of 100 bp  
23166 23873: contig of 708 bp in length  
23874 23973: gap of 100 bp  
23974 24676: contig of 703 bp in length  
24677 24776: gap of 100 bp  
24777 25507: contig of 731 bp in length  
25508 25607: gap of 100 bp  
25608 26297: contig of 690 bp in length  
26298 26397: gap of 100 bp  
26398 27083: contig of 686 bp in length  
27084 27183: gap of 100 bp  
27184 27879: contig of 696 bp in length  
27880 27979: gap of 100 bp  
27980 28674: contig of 695 bp in length  
28675 28774: gap of 100 bp  
28775 29461: contig of 687 bp in length  
29462 29561: gap of 100 bp  
29562 30243: contig of 682 bp in length  
30244 30343: gap of 100 bp  
30344 31064: contig of 721 bp in length  
31065 31164: gap of 100 bp  
31165 31860: contig of 696 bp in length  
31861 31960: gap of 100 bp  
31961 32669: contig of 709 bp in length  
32670 32769: gap of 100 bp  
32770 33480: contig of 711 bp in length  
33481 33580: gap of 100 bp  
33581 34283: contig of 703 bp in length  
34284 34383: gap of 100 bp  
34384 35083: contig of 700 bp in length  
35084 35183: gap of 100 bp  
35184 35907: contig of 724 bp in length  
35908 36007: gap of 100 bp  
36008 36729: contig of 722 bp in length  
36730 36829: gap of 100 bp  
36830 37539: contig of 710 bp in length  
37540 37639: gap of 100 bp  
37640 38334: contig of 695 bp in length  
38335 38434: gap of 100 bp  
38435 39155: contig of 721 bp in length  
39156 39255: gap of 100 bp  
39256 39955: contig of 700 bp in length  
39956 40055: gap of 100 bp  
40056 40792: contig of 737 bp in length  
40793 40892: gap of 100 bp  
40893 41591: contig of 699 bp in length  
41592 41691: gap of 100 bp  
41692 42411: contig of 720 bp in length  
42412 42511: gap of 100 bp  
42512 43231: contig of 720 bp in length  
43232 43331: gap of 100 bp  
43332 44040: contig of 709 bp in length  
44041 44140: gap of 100 bp  
44141 44840: contig of 700 bp in length  
44841 44940: gap of 100 bp



```
* 44941 45640: contig of 700 bp in length
* 45641 45740: gap of 100 bp
* 45741 46442: contig of 702 bp in length
* 46443 46542: gap of 100 bp
* 46543 47250: contig of 708 bp in length
* 47251 47350: gap of 100 bp
* 47351 48061: contig of 711 bp in length
* 48062 48161: gap of 100 bp
* 48162 48853: contig of 692 bp in length
* 48854 48953: gap of 100 bp
* 48954 49680: contig of 727 bp in length
* 49681 49780: gap of 100 bp
* 49781 50484: contig of 704 bp in length
* 50485 50584: gap of 100 bp
* 50585 51294: contig of 710 bp in length
* 51295 51394: gap of 100 bp
* 51395 52092: contig of 698 bp in length
* 52093 52192: gap of 100 bp
* 52193 52884: contig of 692 bp in length.
FEATURES
    source      Location/Qualifiers
                1..52884
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="RP11-46L10"
                /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT    10147 a 13058 c 12120 g 10692 t 6867 others
ORIGIN
Query Match      9.78; Score 52.4; DB 73; Length 52884;
Best Local Similarity 49.18; Pred. No. 1.7;
Matches 131; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 1 cccgcgctgagggcgaggaacgggtataagaagcctcgtggccttgcgcggcagccg 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40493 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40552
QY 61 caggttcccgcgcgccccgagccccgcgcgcctgaagctgcgcgcctctctgggctct 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40553 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40612
QY 121 gcgtgcccctctcgtcagctccgctgctgtcttcttagtggtgcgcgaagcctgtgg 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40613 GCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40672
QY 181 cccagctgtcgtcgtcgtgagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40673 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40732
QY 241 cctcggcacccctcaaccgcgtgaac 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40733 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40759

RESULT 11
LOCUS      AC090977 149719 bp DNA HTG 17-APR-2001
DEFINITION Mus musculus chromosome 16 clone rp23-213m14, WORKING DRAFT
            SEQUENCE, 40 unchromed pieces.
ACCESSION AC090977
VERSION    AC090977.6 GI:13654359
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 149719)
AUTHORS    Milam,J., Ford,B., Hine,R. and Roe,B.A.
TITLE      Mus musculus Chromosome 16 BAC Clone rp23-213m14
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 149719)
AUTHORS     Milam,J., Ford,B., Hine,R. and Roe,B.A.
TITLE      Direct Submission
JOURNAL    Submitted (22-MAR-2001) Department Of Chemistry And Biochemistry,
```

COMMENT

The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

On Apr 17, 2001 this sequence version replaced gi:13592223.

\* NOTE: This is a 'working draft' sequence. It currently consists of 40 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 2175: contig of 2175 bp in length

\* 2176: gap of unknown length

\* 2276: contig of 2132 bp in length

\* 4407: gap of unknown length

\* 4408: contig of 2522 bp in length

\* 7029: gap of unknown length

\* 7123: contig of 2454 bp in length

\* 9583: gap of unknown length

\* 9683: contig of 2278 bp in length

\* 11961: gap of unknown length

\* 12062: contig of 2221 bp in length

\* 14282: gap of unknown length

\* 14283: contig of 2734 bp in length

\* 17116: gap of unknown length

\* 17217: contig of 1980 bp in length

\* 19197: gap of unknown length

\* 19296: contig of 2643 bp in length

\* 21939: gap of unknown length

\* 22040: contig of 2673 bp in length

\* 24712: gap of unknown length

\* 24812: contig of 2623 bp in length

\* 27436: gap of unknown length

\* 27536: contig of 2554 bp in length

\* 30189: gap of unknown length

\* 30190: contig of 2574 bp in length

\* 32863: gap of unknown length

\* 34849: contig of 1986 bp in length

\* 34950: gap of unknown length

\* 37303: contig of 2354 bp in length

\* 37404: gap of unknown length

\* 39634: contig of 2230 bp in length

\* 39734: gap of unknown length

\* 43074: contig of 3340 bp in length

\* 43174: gap of unknown length

\* 45900: contig of 2727 bp in length

\* 45901: gap of unknown length

\* 46001: contig of 3294 bp in length

\* 49295: gap of unknown length

\* 49395: contig of 3531 bp in length

\* 52926: gap of unknown length

\* 53026: contig of 2365 bp in length

\* 55391: gap of unknown length

\* 55491: contig of 3728 bp in length

\* 59319: gap of unknown length

\* 59319: contig of 4147 bp in length

\* 63466: gap of unknown length

\* 63566: contig of 1984 bp in length

\* 65550: gap of unknown length

\* 65650: contig of 4140 bp in length

\* 69790: gap of unknown length

\* 69890: contig of 2687 bp in length

\* 72577: gap of unknown length

\* 72577: contig of 3205 bp in length

\* 75881: gap of unknown length

\* 75882: contig of 3779 bp in length

\* 79761: gap of unknown length

\* 79861: contig of 4790 bp in length

\* 84651: gap of unknown length

\* 84751: contig of 3922 bp in length

\* 88673: gap of unknown length

\* 88772: contig of 4163 bp in length

\* 92936: gap of unknown length

OK 73019, USA

[illegible]





TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL MEDLINE	97000351
REFERENCE	2 (bases 1 to 37931)
AUTHORS	Brown,S.P. and Harris,D.
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 37931)
AUTHORS	Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE	Direct Submission
JOURNAL	Submitted (28-JUN-2000) Streptomycetes coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT	Notes: Streptomycetes coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S-coelicolor/) CDS are numbered using the following system eg SC7B7_01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
FEATURES	Cosmid D10 .. Location/Qualifiers 1..37931 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid D10" 1..786 /gene="SCD10_01" <1..786 /gene="SCD10_01" /notes="SCD10_01, unknown (fragment), len: >261 aa. High content in glycine and aspartate amino acid residues. Contains 2x Pfam match to entry PF01839 FG-GAP, FG-GAP repeat" /codon_start=1 /transl_table=1 /product="hypothetical protein SCD10_01c (fragment)" /protein_id="CAB95880.1" /db_xref="GI:8894719"
gene	
CDS	
misc_feature	

```
/note="Pfam match to entry PF01839 FG-GAP, FG-GAP repeat,
score 19.10, E-value 0.066"
559.729
/misc_feature
/gene="SCD10.01"
/note="Pfam match to entry PF01839 FG-GAP, FG-GAP repeat,
score 29.40, E-value 8.3e-05"
1019.2038
/gene="SCD10.02"
/note="SCD10.02"
1019.2038
/gene="SCD10.02"
/note="SCD10.02, possible transposase, len: 339 aa;
identical to TR:CA871807 (EMBL:AL138662), Streptomyces
coelicolor putative transposase SC8E4A.02, 339 aa"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="CA895881.1"
/db_xref="GI:8894720"
/translation="MGEGLQVRCGLHRLHROEALVVRHNAPLTETGRLRLARCVED
GMPVRAERFQVSHTTASRWARYRQLGVTGMSDRSRHPQRPRTAAAEVHLRL
RREHRIGPLARVCGIAASTAHRILVRHGLPPLAALDRATGEPVRRYERARPGELVH
IDVKILGRIPDGGHKTLGRAGHSRTNGAGWYLHTALDDHSRIAYTEDLPDETAP
TCAAFVLRATAYFASLIGIRIERVLTDNWAYSKNTWRNTCRDLDISPWRTRPWPQTN
GKVERFHTLLEWAYQRPYTSDBERREAFTHLHWYNHRRPHTGIGGHTPASRGTNL
SQQHT"
2125.2128
2134.2697
/gene="SCD10.03"
2134.2697
/gene="SCD10.03"
/note="SCD10.03, conserved hypothetical protein, len: 187
aa; similar to TR:CA893416 (EMBL:AL357524) Streptomyces
coelicolor hypothetical 20.0 kD protein SCD12A.25, 180 aa;
fasta scores: opt: 216 z-score: 277.0 E(): 5.9e-08; 32.0%
identity in 178 aa overlap. Contains also a 3x degenerate
repeat (T/P)TTE"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CA895882.1"
/db_xref="GI:8894721"
/translation="MTWTEPTTEPTTEATTKPLTGERADLLETAKHRHFLFTTRDL
TDEAGLRITVSELCGLGLIKHTAAERNWTFIVNGFSAMGDFTAMTEADWARRADE
FKMLGPEATLVGLDVAEYARTEELIALPDLAGAHLPLKAPFEATWARSARVLL
HVIAETAQHAGHADIIRALDGAKSG"
2137.2172
/note="3x repeat unit translated into SCD10.03"
complement(2741..3355)
/gene="SCD10.04c"
complement(2741..3355)
/gene="SCD10.04c"
/note="SCD10.04c, unknown, len: 204 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCD10.04c"
/protein_id="CA895883.1"
/db_xref="GI:8894722"
/translation="MLFVDSRGELADTFGEHRLATLPASAFPPLAADTEGARLLRTVG
APTGLLRAPDEESGRPLVSRVYVHAEDFAASRDAGWPVIGWLLNAHLALDPGSG
TVYAFDPDEETVRLKLTDSVSLRVYRTARFORLEETFADDEDEAGFERLERAVRY
REETHVDPLPQDDTEVSVVGEIANGQRFAGSPGARSLYG"
complement(3364..3367)
complement(3420..3974)
/gene="SCD10.05c"
complement(3420..3974)
/gene="SCD10.05c"
/note="SCD10.05c, possible mut-like protein, len: 184 aa;
similar to SW:VZGD_BACSU (EMBL:U01168) Bacillus subtilis
hypothetical 45.4 kD protein in thiaminase I 5' region
YgzD, 413 aa; fasta scores: opt: 225 z-score: 279.9 E():
4.1e-08; 35.1% identity in 148 aa overlap. Contains pfam
match to entry PF00293 mutT, Bacterial mutT protein and
match to Prosite entry PS00893 mutT domain signature"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CA895884.1"
/db_xref="GI:8894723"
/translation="MSGCORASVRRAGTARSTPAHPRAPRHAGTAVHVRRLRLRLMR
LLRPVQSRMLMLVNAKFVVGTVGVDRDDEGRVLMKHLRLWPFQWGLPFGFAHGED
FRQTVVREVRETEGLDVEAGRLVNLNGLRLLEVAYEARLLGGELRLDPFETLEARW
CRPDELPEVQVPCPLVRGETVP"
complement(3624..3749)
/gene="SCD10.05c"
/note="Pfam match to entry PF00293 mutT, Bacterial mutT
protein, score 29.20, E-value 7.6e-07"
complement(3633..3692)
/gene="SCD10.05c"
/note="PS00893 mutT domain signature"
4158.5045
/gene="SCD10.06"
4158.5045
/gene="SCD10.06"
/note="SCD10.06, possible integral membrane protein, len:
295 aa; similar to TR:Q9RUT6 (EMBL:AE001976) Deinococcus
radiodurans PecM-related protein DR1295, 302 aa; fasta
scores: opt: 769 z-score: 839.6 E(): 0; 47.1% identity in
291 aa overlap. Contains pfam match to entry PF00892 DUF6,
integral membrane protein. Contains also possible
hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CA895885.1"
/db_xref="GI:8894724"
/translation="MPDGAPGGRFGALGPVGLVLAGGISVQFCAALAVSLMPRAGALG
VVTLLRAVAAVVMLLVCRPLRHGRADWGTVVVFCIAMAGNGLFYQAVDRPLPGA
```

```
Query Match 9.4%; Score 51; DB 3; Length 37931;
Best Local Similarity 45.9%; Pred. No. 3.3;
Matches 174; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 74 cgcgccgaagcccccgcgcgaagctcgccgcctcctgggctctgctggtgcccctgtc 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18060 CGCGCGCGAGACCGCGCGCGGTGTCGCCGCCCTCTACACCGCGCGCT 18119

QY 134 ctgcagctccgctgtgctttcttagtggctcgccaaagcctgtgcccagcctgtcgc 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18120 CTTCCGCGCGCCTGCACCTGTGGTGGCGCCTCCAACGAGGAGCAGCTGCGCGCGC 18179

QY 194 tgcgctggagtcggcgagcgagcgccggggaccctggccaaacccctcggaacctc 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18180 GGTACCGAGCTGGAGCGCGCGCTCGCGCGCGAGAGCCACCGCATCGCGCTCAACTCT 18239

QY 254 caacccgctgaagctcctgctgagcagcctgggcatccccctgtaaacacccctcatagagg 313
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18240 GGGCGCGAGAGACCGCGCGCGCGCGGAAACCGTCCAGGGCTGTCTGCACATGGC 18299

QY 314 ctcccgagaagtgtgtgctgagctgggttccccagggccgtggggcgctgaaggccctgaa 373
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18300 CGCGCGCCTGGGCTGCCCAACCTCTACCGAGCGGGGACGCGGGGACGGGTGGT 18359

QY 374 gacctgtcggggccctgacagtgttgctgagccgagactggagcatctacacctga 433
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18360 GGAGCAGTGGCGCGCGCTGTGACCGCGCGCTGGCACATGACCGCGCTCGCGGACCGG 18419

QY 434 ggacaagacgctgcccacc 452
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18420 GGCACGTGTCGGTGGCGCGC 18438
```

Search completed: November 17, 2001, 15:01:10  
Job time: 8024 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:48:23 ; Search time 115.94 Seconds  
(without alignments)  
1060.700 Million cell updates/sec

Title: US-09-700-770-6

Perfect score: 543

Sequence: 1 ccggcgctggaggcgagg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCRUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	56.4	10.4	263	2	US-08-964-725-2
2	56.4	10.4	507	2	US-08-964-725-4
3	56.4	10.4	519	2	US-08-964-725-5
4	45.6	8.4	6453	1	US-08-306-691B-14
5	45.6	8.4	6453	3	US-09-356-952-8
6	45.4	8.4	2721	6	5215881-2
7	45.4	8.4	8438	1	US-07-945-283-1
8	44.4	8.2	1209	4	US-09-105-537-21
9	44.4	8.2	13613	4	US-09-105-537-3
10	44.4	8.2	38506	3	US-09-320-878-19
11	44	8.1	6453	3	US-09-209-668-10
12	43.6	8.0	432	1	US-08-642-255-48
13	43.6	8.0	756	1	US-08-642-255-50
14	42.8	7.9	13842	4	US-09-105-537-30
15	42.8	7.9	36778	4	US-09-105-537-5
16	42.6	7.8	2543	3	US-08-555-669-11
17	42.6	7.8	2543	3	US-09-073-663-11
18	42.2	7.8	30001	2	US-08-125-468-1
19	42.2	7.8	30001	2	US-08-474-933-1
20	42.2	7.8	4403765	4	US-09-103-840A-2
21	41.8	7.7	4257	2	US-08-690-473-1
22	41.8	7.7	4257	4	US-09-259-821A-1
23	41.8	7.7	4257	4	US-08-843-659-1
24	41.8	7.7	12001	1	US-08-458-568A-11
25	41.6	7.7	530	3	US-08-758-662-4
26	41.6	7.7	4085	1	US-08-486-270-7
27	41.6	7.7	4085	3	US-08-367-264-7

28	41.6	7.7	4181	1	US-08-486-270-9
29	41.6	7.7	4181	3	US-08-367-264-9
30	41.6	7.7	4524	2	US-08-845-998-7
c 31	41.6	7.7	4524	2	US-08-845-998-7
32	41.6	7.7	4524	3	US-09-206-537-7
c 33	41.6	7.7	4524	3	US-09-206-537-7
34	41.6	7.7	4524	4	US-09-430-854-7
c 35	41.6	7.7	4524	4	US-09-430-854-7
36	40.6	7.5	2003	1	US-08-459-526A-21
c 37	40.6	7.5	2003	2	US-08-734-591A-21
38	40.6	7.5	2003	3	US-08-341-018-71
c 39	40.6	7.5	2003	3	US-08-470-335-21
40	40.6	7.5	2003	4	US-08-735-021-21
c 41	40.6	7.5	2003	4	US-08-734-564A-21
42	40.6	7.5	2003	4	US-08-470-339-21
c 43	40.4	7.4	1565	3	US-09-320-878-24
44	40.4	7.4	4195	1	US-08-340-011-1
45	40.4	7.4	4195	3	US-08-901-710-1

ALIGNMENTS

RESULT 1  
US-08-964-725-2  
; Sequence 2, Application US/08964725  
; Patent No. 5939265  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, Maurice  
; APPLICANT: FRIEDMAN, Paula N.  
; APPLICANT: GORDON, Julian  
; APPLICANT: HODGES, Steven C.  
; APPLICANT: KLASS, Michael R.  
; APPLICANT: KRATOCHVIL, Jon D.  
; APPLICANT: ROBERTS-RAPP, Lisa  
; APPLICANT: RUSSELL, John C.  
; APPLICANT: STROUPE, Steven D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE LUNG  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/964,725  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 5997.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 263 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-964-725-2

Query Match 10.4%; Score 56.4; DB 2; Length 263;  
Best Local Similarity 60.4%; Pred. No. 0.00077;  
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 229 ccttggaacaccccttggaacccctcaacccgtgaagctcctgagcagcctgggca 288  
Db 60 CTCTGGACAACATCTTCCCTTTATGGATCCATTAAAGCTTCTTCTGAAAACTCTGGGCA 119  
Qy 289 tccccgtgaacacccctcaatagaggtccccagaagtgtgtgctgagctgggtccccag 348  
Db 120 TTCTCTGTGACACCTGTGTGGAGGGCTTAAGGAAGTGTGTAATAGCTGGGACACGAGG 179  
Qy 349 ccgtggggccgtgaagccctgaagccctgct 382  
Db 180 CTCTGAAGCTGTGAAGAACTGCTGGAGGCGCT 213

## RESULT 2

US-08-964-725-4  
; Sequence 4, Application US/08964725  
; Patent No. 5939265  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, Maurice  
; APPLICANT: FRIEDMAN, Paula N.  
; APPLICANT: GORDON, Julian  
; APPLICANT: HODGES, Steven C.  
; APPLICANT: KLASS, Michael R.  
; APPLICANT: KRATOCHVIL, Jon D.  
; APPLICANT: ROBERTS-RAPP, Lisa  
; APPLICANT: RUSSELL, John C.  
; APPLICANT: STROUPE, Steven D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/964,725  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 5997.US.P1  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 507 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-964-725-4

Query Match 10.4%; Score 56.4; DB 2; Length 507;  
Best Local Similarity 60.4%; Pred. No. 0.00082;  
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 229 ccttggaacaccccttggaacccctcaacccgtgaagctcctgagcagcctgggca 288  
Db 195 CTCTGGACAACATCTTCCCTTTATGGATCCATTAAAGCTTCTTCTGAAAACTCTGGGCA 254  
Qy 289 tccccgtgaacacccctcaatagaggtccccagaagtgtgtgctgagctgggtccccag 348  
Db 255 TTCTCTGTGACACCTGTGTGGAGGGCTTAAGGAAGTGTGTAATAGCTGGGACACGAGG 314  
Qy 349 ccgtggggccgtgaagccctgaagccctgct 382  
Db 315 CTCTGAAGCTGTGAAGAACTGCTGGAGGCGCT 348

## RESULT 3

US-08-964-725-5  
; Sequence 5, Application US/08964725  
; Patent No. 5939265  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, Maurice  
; APPLICANT: FRIEDMAN, Paula N.  
; APPLICANT: GORDON, Julian  
; APPLICANT: HODGES, Steven C.  
; APPLICANT: KLASS, Michael R.  
; APPLICANT: KRATOCHVIL, Jon D.  
; APPLICANT: ROBERTS-RAPP, Lisa  
; APPLICANT: RUSSELL, John C.  
; APPLICANT: STROUPE, Steven D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/964,725  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 5997.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 519 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-964-725-5

Query Match 10.4%; Score 56.4; DB 2; Length 519;  
Best Local Similarity 60.4%; Pred. No. 0.00082;  
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;



Db	2724	ACGGCAGGGCAGTGCAGGAGGCGAGGGCCGGGGTCTGGGCTCACCCCTGCAGTCCGCG	2783
Qy	409	ccgagactggagcatcatcacctgaggacaagacgctgccaccgcgagggctgaaac	468
Db	2784	CCGACACAGCTCCGGGGAAGGGGAGGTCTTTGGGGAGAGCTGGCCCTGAGCCAGGCGCGA	2843
Qy	469	ccgcgcgcggggaggaccgtcccatcccttcccccq	504
Db	2844	GCGGTGACCTCGGGGCCGGGCCCTCTTGTGCCCCAG	2879

RESULT 5  
US-09-356-952-8  
? Sequence 8, Application US/09356952  
? Patent No. 6117663  
? GENERAL INFORMATION:  
? APPLICANT: Boriack-Sjodin, Ann  
? APPLICANT: Margarit, S. M.  
? APPLICANT: Bor-Sogi, Dafna  
? APPLICANT: Cole, Philip  
? APPLICANT: Kuriyan, John  
? TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
? FILE REFERENCE: 600-1-228N  
? CURRENT APPLICATION NUMBER: US/09/356,952  
? CURRENT FILING DATE: 1999-07-19  
? EARLIER APPLICATION NUMBER: 60/093,631  
? EARLIER FILING DATE: 1998-07-21  
? NUMBER OF SEQ ID NOS: 14  
? SOFTWARE: PatentIn Ver. 2.0  
? SEQ ID NO 8  
? LENGTH: 6453  
? TYPE: DNA  
? ORGANISM: Homo sapiens  
US-09-356-952-8

Query Match	8.4%;	Score 45.6;	DB 3;	Length 6453;
Best Local Similarity	47.8%;	Pred. No. 0.18;		
Matches 132;	Conservative 0;	Mismatches 144;	Indels 0;	Gaps

  

Qy	229	ccctggccaaacccctcggcacccctcaaccccgctgaagctctctgctgagcagctctgggca	288
Db	2604	cactgacctctcccttgacacaggcagccgctctgctctagctccagctccgggacc	2663
Qy	289	tcctccgtgaaccacctcatagaggctccccaagatgtgtggctgagctgggtcccacgg	348
Db	2664	ctctgggacccccgggaccctatgtaccaccagcgccctctgcactgtaggtctcccggg	2723
Qy	349	ccgtggggggcgtgaagggccctgaagccctgctctggggccctgcagctgtttggctgag	408
Db	2724	acggcagggcagtgaggaggcgagggcgggctctgggctcacgcccctgcagtcctggg	2783
Qy	409	ccagagctggagcatctacacctgagggacaaagacgctgcccacccgagagggtgaaac	468
Db	2784	ccgacacagctccgggaaaggcgaggtctctggggagagctgcctgagccagccgga	2843
Qy	469	ccgcgcgcggggaggaccgtccatccccctcccccg	504
Db	2844	gcggtgacctggggccgcctcttgcctccag	2879

  

RESULT	6
5215881-2/c	
; Patent No. 5215881	
; APPLICANT: CHEUNG, ANDREW K.	
; TITLE OF INVENTION: PSEUDORABIES DIAGNOSIS PROBES	
; NUMBER OF SEQUENCES: 3	
; CURRENT APPLICATION DATA:	
; APPLICATION NUMBER: US/07/537,855	
; FILING DATE: 13-JUN-1990	
; FILING DATE:	









**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 13:34:21 ; Search time 2570.26 Seconds  
(without alignments)  
1997.035 Million cell updates/sec

Title: US-09-700-770-6  
Perfect score: 543  
Sequence: 1 ccggcgctgaggggcgagg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues  
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_estl1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estov1:\*  
70: em\_estov2:\*  
71: em\_estov3:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.





## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -400P from Gibco.

## FEATURES

Location/Qualifiers  
1..439  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2248841"  
/clone\_lib="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

71 a 140 c 155 g 73 t

## Query Match

Best Local Similarity 80.9%; Score 439; DB 23; Length 439;  
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 cgcgcctcctgggctctgcgtggccctcctgcagctccgctgctgttcttagtg 162  
Db 439 CGGCGCTCTGGGCTCTGGTGGCCCTGCTCCACCTCGCTCCGCTGCTTCTTAGTGG 380  
QY 163 gctggccagcctgtgtggccagcctgtcgtcgtggagtcggcgagggccggg 222  
Db 379 GCTGGCAAGCCTGTGGCCAGCCTGCTGCTGGTGGAGTCGGCGCGAGCGCGGG 320  
QY 223 ccggaccctggccaaaccctcggaccctcaaccctgaagctcctcgtcagagcc 282  
Db 319 CCGGACCTTGGCCAAACCCCTCGGCACCTCAACCCGCTGAAGCTCCTGTGAGCAGCC 260  
QY 283 tgggcatcccccgtgaaccactcatagagggctcccagaagtgctggcgtgagctgggtc 342  
Db 259 TGCGATCCCGTGAACCACTCATAGAGGGCTCCAGAGTGTGTGGCTGAGCTGGGTC 200  
QY 343 ccaaggcgtggggcgctgaaggccctgaagccctgtcgtggggccctgacagtgttg 402  
Db 199 CCAGGCGGTGGGGCGCTGAAGGCCCTGAAGGCCCTGCTGGGGGCCCTGACAGTGTG 140  
QY 403 gctgagccagactggagatctacacctgagacaagcgtgccaccgcgagggct 462  
Db 139 GCTGAGCCGAGACTGAGCATCTACACCTGAGGAGGCTCCAGAGTGTGTGGCTGAGCTGGGCT 80  
QY 463 gaaaccccgccggggaggaccgtccatcccttcccccgccctctcaataaacgt 522  
Db 79 GAAACCCCGCCGGGAGGAGCGCTCCATCCCTTCCCGCGGCCCTCTCAATAACGT 20  
QY 523 ggttaagacgcaaaaaa 541  
Db 19 GGTAAAGACGCAAAAAA 1

## RESULT

3  
AI187341/c 425 bp mRNA EST 29-OCT-1998  
LOCUS

## DEFINITION

qe26d06.x1 Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
IMAGE:1740107 3', mRNA sequence.

## ACCESSION

AI187341

## VERSION

AI187341.1 GI:3737979

## KEYWORDS

ESF.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 425)

## AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/hiclogap>.

## TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

## JOURNAL

Tumor Gene Index

## COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Insert Length: 532 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 372.

## FEATURES

Location/Qualifiers

1..425

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1740107"

/clone\_lib="Soares\_fetal\_lung\_NbHL19W"

/dev\_stage="19 weeks"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: lung; Vector: pT73D (Pharmacia) with a

modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTT-3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo. This library was constructed

from the same fetus as the fetal heart library, Soares

fetal heart NbHL19W."

BASE COUNT 70 a 139 c 149 g 67 t

## ORIGIN

Query Match 78.3%; Score 425; DB 17; Length 425;  
Best Local Similarity 100.0%; Pred. No. 1.2e-81;  
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ctggggctctgcgtggccctgtcctcagctccgctcgtcgttcttagtggtcgcc 170

Db 425 CTGGGGCTCTGGCGTGGCCCTGTCTCGAGCTCGCTTCTTAGTGGCTGGGC 366

QY 171 aagcctgtggccagcctgtcgtcgtggagtcggcgagagccggccagggacc 230

Db 365 AAGCCTGTGGCCAGCCTGTCTGCTGGCTGGAGTCGGCGCGGAGCCCGGGGACC 306

QY 231 ctggccaaacccctcgtggccctcaaccctgaagctcctcgtgagcagcctgggcatc 290

Db 305 CTGGCCAAACCCCTCGGCACCCCTCAACCCGCTGAAGCTCCTGCTGAGCAGCCTGGGCATC 246

QY 291 ccctgaaaccactcatagagggctcccagaagtgtgtgctgagctgggtcccaaggcc 350

Db 245 CCGGTGAACCACTCATAGAGGCTCCCAAGAGTGTGTGGCTGAGTGGGTCCCGAGGCC 186

QY 351 gtggggccgtgaaaggccctgaagccctgctggggccctgacagtgttggctgagcc 410

Db 185 GTGGGGCGCGTGAAGGCCCTGAAGCCCTGCTGGGGCCCTGACAGCTGTTGGCTGAGCC 126

QY 411 gagactgagacatctacacctgagacaagacgctgcccaccgcgagggctgaaaaacc 470

Db 125 GAGACTGGAGCATCTACACCTGAGGACAGACGCTGCCACCCCGGAGGCTGAAAAACC 66

```

QY 471 cgccgggggagaccgtccatccctcccccggccctctcaataaaacgtgttaaga 530
|||||
Db 65 CGCCGGGGAGGAGCCCTCCATCCCTTCCCGCGGCCCTCTCAATAAACGTGGTTAAGA 6
QY 531 gcaaa 535
|||||
Db 5 GCAA 1

RESULT 4
AI333740/c
LOCUS AI333740.1 424 bp mRNA EST 28-DEC-1998
DEFINITION gp98f05.x1 Soares_fetal_lung_NbHL19w Homo sapiens cDNA clone
IMAGE:1931073 3', mRNA sequence.
ACCESSION AI333740
VERSION AI333740.1 GI:4070299
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 413.

FEATURES
     source
     1..424
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone_lib="Soares_fetal_lung_NbHL19w"
     /dev_stage="19 weeks"
     /lab_host="DH10B (ampicillin resistant)"
     /note="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGAGCGCGCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19w."
BASE COUNT 70 a 139 c 149 g 66 t
ORIGIN

Query Match 78.1%; Score 424; DB 19; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.9e-81;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ctggggctctgcgtggccctgtcctgcagctcccgctgctgtcttcttagtggtcgcc 170
|||||
Db 424 CTGGGGCTCTGCGGGCCCTGCTCTGCAGCTCCGCTGCTGCTTTCTTAGTGGGCTCGGCC 365
QY 171 aagcctgtgcccagcctgtcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 230
|||||
Db 364 AAGCCTGTGCGCCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 305
QY 231 ctggccaaacccctcgacacccctcaacccctgaagctcctgctgagcagcctggcctc 290
|||||
Db 304 CTGGCCAAACCCCTCGGACCCCTCAACCCCTGAAGCTCTCTGCTGAGCAGCCTGGGCATC 245
QY 291 ccgctgaaccacacctagagggtcccccagaagtgtgtggtgagctggtgtcccccaggcc 350

```

```

Db 244 CCCGTGAACCACTCATAGAGGCTCCAGAAAGTGTGTGGCTGAGCTGGGTCCCGAGCC 185
QY 351 gtgggggcccgtgaagccctgaagccctgctggggccctgacagtgatttggtgagcc 410
|||||
Db 184 GTGGGGCCCTGGAAGGCCCTCAAGGCCCTGCTGGGGGCCCTGACAGTGTGTTGGGTGAGCC 125
QY 411 ggaactgagcatctacactgagacgaacgctgccaccgccgagggctgaaaaccc 470
|||||
Db 124 GAGACTGGAGCATCTACACCTGAGACAAGACGCTGCCACCCCGGAGGGCTGAAAACCC 65
QY 471 cgccgggggagaccgtccatccctcccccggccctctcaataaaacgtgttaaga 530
|||||
Db 64 CGCCGGGGGAGGAGCCCTCCATCCCTTCCCGCGGCCCTCTCAATAAACGTGGTTAAGA 5
QY 531 gcaa 534
|||||
Db 4 GCAA 1

RESULT 5
AI744099/c
LOCUS AI744099.1 424 bp mRNA EST 17-DEC-1999
DEFINITION wc36c10.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2317266 3',
mRNA sequence.
ACCESSION AI744099
VERSION AI744099.1 GI:5112387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 495 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES
     source
     1..424
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone_lib="NCI_CGAP_Pr28"
     /sex="male"
     /dev_stage="adult"
     /lab_host="DH10B"
     /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 68 a 133 c 146 g 77 t
ORIGIN

Query Match 77.5%; Score 420.8; DB 24; Length 424;
Best Local Similarity 99.5%; Pred. No. 9.4e-81;

```



Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bhrp/image/image.html  
Insert length: 497 Std Error: 0.00  
Seq primer: -400p from Gibco.

#### FEATURES

source  
1. .413  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2317080"  
/clone\_lib="NCI-CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI-CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneIDs  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 68 a 135 c 146 g 64 t

Query Match 76.1%; Score 413; DB 24; Length 413;

Best Local Similarity 100.0%; Pred. No. 4.4e-79;  
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 tgcgtggccctgtcctgcagctccgctgctgtcttcttagtggctcgcccaagcctgtg 179  
Db 413 TGCCTGGCCCTGTCTGCAGCTCCGCTGCTTTCTTAGTGGCTCGGCCAAGCCTGTG 354

Qy 180 gccagcctgtcgtgcgtgagtcgctggcgagcgccggcgccgacctggccaac 239  
Db 353 GCCCAGCCTGTCTGGCTGGCTGAGTCGGCGCGGAGCGCGGGCGGACCTGGCCAA 294

Qy 240 cccctcgacacctcaacccgtgaagctcgtgctgagcagcctggcgccatcccgtaga 299  
Db 293 CCCCTCGCACCTCAACCCGCTGAGCTCTGCTGAGCAGCCTGGGATCCCGTGAAC 234

Qy 300 cactcatagaggctcccaagaagtgtgtgctgagctgggttccccagggcctggggcc 359  
Db 233 CACCTCATAGAGGCTCCAGAGTGTGTGCTGAGCTGGGTCCCCAGGCGTGGGGCC 174

Qy 360 gtgaagccctgaagccctgctggggccctgacagtgtttgtgctgagcagagactgga 419  
Db 173 GTGAAGCCCTGAAGCCCTCTGGGGCCCTGACAGTGTGTGCTGAGCGAGACTGGA 114

Qy 420 gcatctacacctgaggacaagactgctgccaccgcgagggctgaaaccccgccgcgg 479  
Db 113 GCATCTACACCTGAGGACAAGACGCTGCCACCGCGAGGGCTGAAAACCCCGCGGG 54

Qy 480 gagacgctccatcccttccccggccctctcaataaactgtggttaagagc 532

Db 53 GAGGACCGTCCATCCCTTCCCGGGCCCTCTCAATAAAGTGTTAAGAGC 1

#### RESULT

8  
BF221778/c 416 bp mRNA EST 09-NOV-2000  
LOCUS 7062e02.x1 NCI-CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:3578595 3'  
DEFINITION similar to TR:Q9VI31 Q9VI31 CG10690 PROTEIN. ;, mRNA sequence.  
ACCESSION BF221778  
VERSION BF221778.1 GI:11128955  
KEYWORDS EST.  
SOURCE human.

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

1 (Bases 1 to 416)

#### AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

#### TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

#### JOURNAL

Tumor Gene Index

#### COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

[info@image.llnl.gov](mailto:info@image.llnl.gov)

High quality sequence stop: 383.

#### FEATURES

Location/Qualifiers

source

1. .416

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="3578595"

/clone\_lib="NCI-CGAP\_Pr28"

/sex="male"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)

with a modified polylinker; Plasmid DNA from the

normalized library NCI-CGAP\_Pr22 was prepared, and ss

circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from a pool

of 5,000 clones made from the same library (cloneIDs

985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 69 a 134 c 144 g 69 t

#### ORIGIN

Query Match 76.0%; Score 412.8; DB 146; Length 416;

Best Local Similarity 99.5%; Pred. No. 4.9e-79;

Matches 414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 122 cgtggccctgtcctgcagctccgctgcttcttagtggctcgcccaagcctgtgac 181  
Db 416 CGTGGCCCTGTCTGCAGCTCCGCTGCTTCTTAGTGGCTCGGCCAAGCCTGTGGC 357

Qy 182 ccagcctgtcgtcgctggagtcggtcggtggcgagcgccggcgccctggcccaacc 241  
Db 356 CCAGCCTGTCTGGCTGGAGTCGGCGCGGAGCGCGGGCGGACCTGGCCACCC 297

Qy 242 cctcggcaacctcaaccgctgaagctcctctgctgagcagcctgggcatccccgtgaacca 301  
Db 296 CCTCGCACCTCAACCGTTTGAAGCTCCTTGTAGCAGCCTGGGCATCCCGCTGAACA 237

Qy 302 cctcatagaggtctcccaagaagtgtgtgctgagctaggtcccccagccgtggggccgt 361  
Db 236 CCTCATAGAGGCTCCCAAGAAGTGTGTGCTGAGCTGGGTCCCCAGGCCGTGGGGCCGT 177

Qy 362 gaaggccctgaagccctgctggggccctgacagtgtttggctgagcagagctggagc 421  
Db 176 GAAGCCCTGAAGCCCTGCTGGGGCCCTGACAGTGTTTGGCTGAGCCGAGACTGGAGC 117

Qy 422 atctacacctgagacagacgctgccaccgcgagggctgaaacccccccgcgggga 481  
Db 116 ATCTACACCTGAGGACAAGACGCTGCCACCCCGAGGGCTGAAACCCCGCCGGGGA 57

Qy 482 ggaacgttcattccccctccccggccctctcaataaactgtggttaagagcaaaa 537  
Db 56 GGACCGTCCATCCCTTCCCCCGCCCTCTCAATAAACGTTGTTAAGAGCAAAA 1

```
RESULT 9
AI936195/c 420 bp mRNA EST 06-SEP-1999
LOCUS w63a10.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:245994 3',
DEFINITION mRNA sequence.
ACCESSION AI936195
VERSION AI936195.1 GI:5675065
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 420)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Cloned distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
FEATURES
Source
1..420
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:245994"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
63 a 130 c 143 g 84 t
BASE COUNT 63 a 130 c 143 g 84 t
ORIGIN
Query Match 75.4%; Score 409.4; DB 103; Length 420;
Best Local Similarity 99.8%; Pred. No. 2.6e-78;
Matches 410; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 133 cc:gcagctcgctgctgtttcttagtggtcgccagcctgtggccagcctgtcg 192
Db 420 CCYGCAGCTCCGCTGCTGCTTCTTAGTGGCTCGGCCAGCTGTGGCCAGCCGTG 361
QY 193 ctgcgtgagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 252
Db 360 CTYCGCTGAGTCCGGCGGAGCGCGGGCCGCCGCCCTGGCCAAACCCCTCGGCACCC 301
QY 253 tcacccgcgtgaagctcctgctgagcagcctggtgcatccctgaacacccctcatagag 312
Db 300 TCACCCGCTGAAGCTCTCTGCTGAGCAGCTGGGCATCCCGTGAACCACTCATAGAGG 241
QY 313 gctcccaagtgctgtggtgagctgggttcccaagccctggtgggcgcgtgaaggccctga 372
Db 240 GCYCCAGAAAGTGTGTGCTGAGCTGGGTCCCGAGCCGCTGGGGCCGTGAAGGCCCTGA 181
QY 373 agccctgctggggccctgacagtgcttggctgagccgagactgagacatctacacctg 432
Db 180 AGCCCTGCTGGGGGCCCTGACAGTGTGTTGGTGAAGCCGAGAGACTGAGCATCTACACCTG 121
```

```
QY 433 aggacaagacgtgccaccgcgcagaggtgaaacccgcgcgcgcgcgcgcgcgcgc 492
Db 120 AGACAACACAGCTGCCACCCGCGAGGGCTGAACACCCCGCGGGAGAGCGTCCAT 61
QY 493 cccctccccgcgcgcctctcaataaacgtggttaagagcaaaaaaa 543
Db 60 CCCCTTCCCGCGCCCTCCCAATAAACGTGTTAGAGCAAAAAA 10
RESULT 10
AI684641/c 378 bp mRNA EST 07-MAR-2000
LOCUS wa84h08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2302911 3', mRNA sequence.
ACCESSION AI684641
VERSION AI684641.1 GI:4895935
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 378)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 474 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
Source
1..378
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2302911"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
55 a 125 c 132 g 66 t
BASE COUNT 55 a 125 c 132 g 66 t
ORIGIN
Query Match 69.6%; Score 378; DB 23; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.5e-71;
Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 159 qtgagctcgcccaagcctgtgcccagcctgtcgtgcgtgagtcgagcgagagcc 218
Db 378 GTGGGCTCGGCCAAGCCCTGTGCCCCAGCCTGTGCTGGCTGGAGTCGGCGCGAGGCC 319
QY 219 ggggcgcgcgcctggccaaacccctcgcaacctcaaccgcgtgaagctcctgctgagc 278
Db 318 GGGCGCGGACCTTGGCCAAACCCCTCGGCACCCCTCAACCCGCTGAAGCTCTGCTGAGC 259
QY 279 agcctggcgcctcccgctgaaccacccctcatagaggctcccaagaagtgtgtggtgagctg 338
Db 258 AGCCTGGGCATCCCGTGAACCAACCTCATAGAGGCTCCACAGAGTGTGTGCTGAGCTG 199
QY 339 ggtcccccagcgcgtgggcgcgtgagccctgaaggccctgctgtgggcctgacagtg 398
```

```

Db 198 GGTCCCGAGGCGGTGGGGCGGTGAAGGCCCTGAAGGCCCTGTGGGGCGCCTGACAGTG 139
      |||
Qy 399 ttggctgagccagactgagcatctacacctgagagacagctgccaccgcgag 458
      |||
Db 138 TTTGGCTGAGCCGAGAGCTGAGGACATCTACACCTGAGGACAAAGACGCTGCCACCCGCGAG 79
      |||
Qy 459 ggctgaaacccgcgcggggaggaccgtccatcccttccccgcctctcaataa 518
      |||
Db 78 GGCTGAAACCCCGCGCGGGGAGGACCGTCCATCCCTTCCCCGCGCCCTCTCATAA 19
      |||
Qy 519 acgtggttaagagcaaaa 536
      |||
Db 18 ACCTGGTTAAGAGCAAAA 1
      |||

RESULT 11
AA649864/c
LOCUS
DEFINITION ns54a03.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1187404 3',
      mRNA sequence.
ACCESSION AA649864
VERSION AA649864.1 GI:2577192
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
      Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-femail.nih.gov
      Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
      Emmert-Buck, M.D., Ph.D.
      cDNA Library Preparation: M. Bento Soares, Ph.D.
      DNA Sequencing by: Greg Lennon, Ph.D.
      Cloned through the I.M.A.G.E. Consortium/LLNL at:
      www-bio.llnl.gov/bbrp/image/image.html
      Insert Length: 539 Std Error: 0.00
      Seq primer: -40ml3 fwd. ET from Amersham
      High quality sequence stop: 315.
      Location/Qualifiers
      1. 343
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="IMAGE:1187404"
         /clone_lib="NCI_CGAP_Pr22"
         /sex="male"
         /tissue_type="normal prostate"
         /lab_host="Dhl0B"
         /notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
         with a modified polylinker; 1st strand cDNA was prepared
         from normal prostate bulk tissue, and was then primed with
         a Not I - oligo(dT) primer. Double-stranded cDNA was
         ligated to Eco RI adaptors (Pharmacia), digested with Not
         I and cloned into the Not I and Eco RI sites of the
         modified pT73 vector. Library is normalized, and was
         constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 48 a 111 c 119 g 65 t
ORIGIN

Query Match 62.6%; Score 339.8; DB 10; Length 343;
Best Local Similarity 99.4%; Pred. No. 2.4e-63;
Matches 341; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 195 gcgtgagtcgcgcggagcgccggggaccctggccacccctcgccacctc 254
      |||

```

```

Db 343 GCCTGGAGTGGCGGCGAGCGGGCGGGACCCCTGGCCAAACCCCTCGGCACCCCTC 284
Qy 255 aaccgcctgaagctccctgctgagcagcctgggcacccctgaaccacctcatagaggc 314
      |||
Db 283 AACACGCTGAAGCTCCCTGCTGAGCAGCGCTGGCATCCCGTGAACACCTCATAGAGGC 224
Qy 315 tcccaagaagtgtgtgctgagctgagtggttccccagcgcctgggggcctgaaaggccctgaag 374
      |||
Db 223 TCCCAGAAAGTGTGTGGCTGAGCTGGGTGCCAGGCCCTGGGGCGCTGAAGGCCCTGAAG 164
Qy 375 gccctgctgggggcccctgacagtggttggctgagccgagactgagcatctacacctgag 434
      |||
Db 163 GCCTGCTGGGGGCCCTGACAGTGTGTTGGCTGAGCCGAGACTGGAGCATCTACACCTGAG 104
Qy 435 gacaaagcctgcccaccgcgagggctgaaaccccccgcgcgcggggagagccgctccatcc 494
      |||
Db 103 GACAAGACGCTGCCACCCCGGAGGCTGAAACCCCGCGGGGAGGACCGTCCATCC 44
Qy 495 cttccccccggccctctcaataaacgtggttaagagcaaaaaa 537
      |||
Db 43 CCTTCCCCGCGCCCTCTCAATAAACGTGTTAAGAGCAAAAA 1

RESULT 12
AA902200/c
LOCUS
DEFINITION AA902200 335 bp mRNA EST 09-JUN-1998
      OK69a06.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1519186 3',
      mRNA sequence.
ACCESSION AA902200
VERSION AA902200.1 GI:3037390
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 335)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
      Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-femail.nih.gov
      Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
      Emmert-Buck, M.D., Ph.D.
      cDNA Library Preparation: M. Bento Soares, Ph.D.
      DNA Sequencing by: Washington University Genome Sequencing Center
      Cloned through the I.M.A.G.E. Consortium/LLNL at:
      www-bio.llnl.gov/bbrp/image/image.html
      Insert Length: 600 Std Error: 0.00
      Seq primer: -40ml3 fwd. ET from Amersham
      High quality sequence stop: 297.
      Location/Qualifiers
      1. 335
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="IMAGE:1519186"
         /clone_lib="NCI_CGAP_GC4"
         /tissue_type="pooled germ cell tumors"
         /lab_host="Dhl0B"
         /notes="Vector: pT73D-Pac (Pharmacia) with a modified
         polylinker; 1st strand cDNA was prepared from 3 pooled
         germ cell tumors, and was then primed with a Not I -
         oligo(dT) primer. Double-stranded cDNA was ligated to Eco
         RI adaptors (Pharmacia), digested with Not I and cloned
         into the Not I and Eco RI sites of the modified pT73
         vector. Library is normalized. Library was constructed by
         Bento Soares and M. Fatima Bonaldo."

BASE COUNT 49 a 107 c 120 g 59 t
ORIGIN

```





Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV3-HT0544-190>  
600-225-h06&t3=2000-06-19&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 323.

## FEATURES

```
1. 323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0544"
/dev_stage="Adult"
/note="Organ: head_neck; Site:2; Smal; A mini-lib derived from ORESTES PC No. 196,716 - Ludwig In profiles into the pUC 1 tissue mRNA and cDNA amp low stringency condition
```

Query Match	52.6%	Score	285.4;	DB	139;	Length	323;
Best Local Similarity	96.4%;	Pred.	No. 1.le-51;				
Matches	292;	Conservative	0;	Mismatches	11;	Indels	0;
Gaps	0;						
Qy	204	tcgcgcggaaggccggggcggagaccctggccaacccccctcgcacctcaaccgcgtg	263				
Db	20	TCGGGGGCGCTTTCGTGCGCGGGACCTTGCCAAACCCCCTCGGCACCCTCAACCCGCTG	79				
Qy	264	aagctctctgtgagcagcctgggcatcccgctgaaccacctcatagaggctcccagaag	323				
Db	80	AAGTCTCTGTGACGACCTGGGCATCCCCGTGAACACCCTCATAGAGGGCTCCAGAAG	139				
Qy	324	tgtgtgactgaagtgggtccccagggcgtggggcgtgaaagccctgaagccctgctg	383				
Db	140	TGTGTGCTGAGCTGGGTTGCCAGCGCGTGGGGCCGTGAAGGGCCTGAAGGGCCTGTGTG	199				
Qy	384	ggggccctgacagtgttggctgagcgcaactggacatcataccttgagacaagacg	443				
Db	200	GGGECCTTGACAGTGTTTGGCTGAGCCGAGACTGGAGCATCTACACCTTGAGGACAAGACG	259				
Qy	444	ctgccaccgcgagggtgaaaacccccgcgcggggaggagccgtccatccccctcccc	503				
Db	260	CTGCCACCCGCGAGGGCTGAAAACCCCGCGGGGAGGACGTCCATCCCCCTCCCCA	319				
Qy	504	ggc	506				
Db	320	CGC	322				

RESULT 15  
W72786/C

W72786	389 bp	mRNA	EST	16-OCT-1996
zd61a04.s1	Soares_fetal_heart	NbHH19W	Homo sapiens	clone
IMAGE:345102	3'	mRNA sequence.		

ACCESSION W72786  
VERSION W72786.1 GI:1382821  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria;  
 1 (bases 1 to 389)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,  
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
 Rifkin, L., Rohlff, T., Soares, N., Tan, F., Traversakis, E., Waterston

TITLE	JOURNAL	COMMENT

R., Williamson, A., Wohldmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL & con-  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further infor-  
Insert Length: 531 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 256.

FEATURES	SOURCE
1. <i>Age</i>	1. <i>Age</i>
2. <i>Gender</i>	2. <i>Gender</i>
3. <i>Marital Status</i>	3. <i>Marital Status</i>
4. <i>Education</i>	4. <i>Education</i>
5. <i>Occupation</i>	5. <i>Occupation</i>
6. <i>Income</i>	6. <i>Income</i>
7. <i>Religion</i>	7. <i>Religion</i>
8. <i>Political Party</i>	8. <i>Political Party</i>
9. <i>Health Status</i>	9. <i>Health Status</i>
10. <i>Family Size</i>	10. <i>Family Size</i>
11. <i>Home Ownership</i>	11. <i>Home Ownership</i>
12. <i>Travel Frequency</i>	12. <i>Travel Frequency</i>
13. <i>Spending Habits</i>	13. <i>Spending Habits</i>
14. <i>Life Satisfaction</i>	14. <i>Life Satisfaction</i>
15. <i>Community Involvement</i>	15. <i>Community Involvement</i>
16. <i>Employment Status</i>	16. <i>Employment Status</i>
17. <i>Health Insurance</i>	17. <i>Health Insurance</i>
18. <i>Retirement Savings</i>	18. <i>Retirement Savings</i>
19. <i>Charitable Giving</i>	19. <i>Charitable Giving</i>
20. <i>Life Expectancy</i>	20. <i>Life Expectancy</i>

```

location/Vectors
1. .389
/organism="Homo sapiens"
/db_xref="GDB:1270477"
/db_xref="taxon:9606"
/clone="IMAGE:345102"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TCTTACCAATCTGAAGTGGAGCGGCATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHH19W."

```

BASE COUNT		NONL19W.	
ORIGIN			
64 a	124 c	120 g	67 t
14 others			

Query Match	Score	DB	Length
Best Local Similarity	51.6%	280;	190;
Matches	91.3%	Pred. No. 1.7e-50;	
Matches 335: Conservative	0;	Mismatches 26;	
Matches 335: Indels	0;	Mismatches 26;	
Matches 335: Gaps	0;	Mismatches 26;	

Qy	171	aagcctgtggccc--agcctgtcgtcgtcgtgagtcgg--cgcgagagccggagccggcg	227
Db	365	AAGCCTTGGCCNAAGCCTTNGTTCTGNGCTGGAGTCGGCGGNGGTAGCGCGGGCGCGG	306
Qy	228	acctggccaaacccctctggcaccctcaaccgcctgaagcctcctcgtgagcagcctggcg	287
Db	305	ACCTTGCNAAACCCCTCGGNACCCTCAACCCGCTGAAGCTCCTGCTGAGCAGCCTTGGCG	246
Qy	288	atccccctgaacc--acctatagagggctcccagaagctgtgctggaactcgtgggtcccca	346
Db	245	ATCCCCGTGAACCNACCTCATAGAGGGCTCCCCAGAAGTGTGGCTTGAGCTGGGTCCCCA	186
Qy	347	ggccgtggggcgcgtaagggccctgaagccctgctgggggcccctgacagtggtttggctg	406
Db	185	GGCCGTGGGG--CCGTGAACCCCTGNAGGCCTGCTTGGGGCCCTCACAGTGTTTGGCTG	128
Qy	407	agccgagactggagcatctacacctgaggacaagcgtgccaccgccgagggctgaaa	466
Db	127	AGCCGAGACTGGAGCATCTACACTGAGGANAAGACGCTGCCACCCTCGAGGGNTGAA	68
Qy	467	accgcgcgcgggagacogtccatccctctcccccgcctctccataaacctggtt	526
Db	67	TACCCGCGGGGAGACCGTCATCCCTTCCCGCGGCNNCTTCATAAACCTGGTT	8

Search completed: November 17, 2001, 13:34:25  
Job time: 3424 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:53:48 ; Search time 4216.79 Seconds  
(without alignments)  
7556.355 Million cell updates/sec

Title: US-09-700-770-3  
Perfect score: 2060  
Sequence: 1 cttgagagctctcaataact.....ttcattgaaaaa

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_on:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: gb\_ba1:\*  
17: gb\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_on:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v1:\*  
59: gb\_v2:\*  
60: gb\_vtg1:\*  
61: gb\_vtg2:\*  
62: gb\_vtg3:\*  
63: gb\_vtg4:\*  
64: gb\_vtg5:\*  
65: gb\_vtg6:\*  
66: gb\_vtg7:\*  
67: gb\_vtg8:\*  
68: gb\_vtg9:\*  
69: gb\_vtg10:\*  
70: gb\_vtg11:\*  
71: gb\_vtg12:\*  
72: gb\_vtg13:\*  
73: gb\_vtg14:\*  
74: gb\_vtg15:\*  
75: gb\_vtg16:\*  
76: gb\_vtg17:\*  
77: gb\_vtg18:\*  
78: gb\_vtg19:\*  
79: gb\_vtg20:\*  
80: gb\_vtg21:\*  
81: gb\_vtg22:\*  
82: gb\_vtg23:\*  
83: gb\_vtg24:\*  
84: gb\_vtg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1358.4	65.9	1438	88	AF098484	AF098484 Homo sapi
2	1351.8	65.6	1358	89	AF200345	AF200345 Homo sapi
3	1350.4	65.6	1519	10	E33553	E33553 ASP5. 2/200
4	1350.4	65.6	1648	10	E33552	E33552 ASP5. 2/200
5	1345.2	65.3	1365	88	AF090386	AF090386 Homo sapi
6	1199.2	58.2	1299	9	AR016588	AR016588 Sequence
7	1199.2	58.2	1299	9	AR078758	AR078758 Sequence
8	1151.8	55.9	1375	9	A70198	A70198 Sequence 2





Db	781	ASATCCACATGGAGCGTGAAGTGGCCCGAGGCTGACTCTGTGCCAAGGCTGTG	840	
Qy	1546	c-gccatcctggatacagggcacgtccctcatcacagagaccactgaggagatccggccc	1605	
Db	841	CTGCCATCTTGGATACAGGCACGTCCTCATCATCAGAGCCACTGAGGAGATCCGGGCC	900	
Qy	1606	tjcatgacgacattgggggaatccctctgtgctgagtgaggagatcatcatctgtctcg	1665	
Db	901	TSCATGAGGCATTTGGGGGAATCCCTTTGCTGGCTGGGGAGTACATCATCTGTGTCGG	960	
Qy	1666	aaatcccaaaagctcccgacgtctccttctcttgggggggtctggtttaacctcaacg	1725	
Db	961	AAATCCCAAGCTCCCGCAGTCTCTTCTCTTGGGGGGTCTGTGTTAACTCAGCG	1020	
Qy	1726	cccatgattacgtcatccagactactcgaatggcgctccgctctgtctgtccgggttcc	1785	
Db	1021	CCCATGATTACGTATCCAGACTACTCGAAATGGCGTCCGCCTCTGTCTTCCGGTTCC	1080	
Qy	1786	agccctggatgtccctccgctcagggcccttctgtgacctcggtgacgtctcttgg	1845	
Db	1081	A3GCCCTGGAGTCCCTCCGCTCGAGGCCCTTCTGGATCCCTCGGTGACGTCTCTTGG	1140	
Qy	1846	ggacgtatgtgcccgtcttcgaccgggggacatgaagagcagcccggtgggacctgg	1905	
Db	1141	GGACGTATGTGGCGCTCTTCGACCGCGGACATGAAGAGCAGCGCCCGGTGGCGCTGG	1200	
Qy	1906	qgcgcgtgcacitcgggagagacactcgatgggagagactgcgacgycagttcc	1965	
Db	1201	CGCGCGTCCGACATCGCGGAGCGACCTCGGATGGGAGAGACTGCCGAGCGCAGTCC	1260	
Qy	1966	ccgggtgaccccgaatgaacgcatcgacgcggtggtcgagaggtctcgtacccca	2025	
Db	1261	CGGGTGACGCCCAAGTGAACGCGATCGCAGCGGTTGGTGGCGAGGCTCTGTACCCA	1320	
Qy	2026	gtaaaaatccactatttccattgataaaaaa	2060	
Db	1321	GTAAAAATCCACTATTTCATTGTAATAAAAAA	1355	
RESULT	3			
E33553				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
DAVID J POWER, JOHN KEI, JEFFREY HILL, TUDY SMITH PC				
C12N15/09, A61K38/00, A61K39/395, A61K45/00, A61K48/00, PC				
C07K16/18,				
PC C12N5/10, C12P21/02, G01N33/53, G01N33/566, C12N15/00, A61K37/02,				
CC C12N5/00				
CC				
FH				
FT				
FT				
FEATURES				
SOURCE				

BASE COUNT	304	a	446	c	436	g	331	t	2	others
ORIGIN										
Query Match	65.6%; Score 1350.4; DB 10; Length 1519;									
Best Local Similarity	99.6%; Pred. No. 0;									
Matches 1354; Conservative	0;	Mismatches	6;	Indels	0;	Gaps	0;			
Qy	701	gtccccagcagatgtctccaccacccgctgctgcaacccctgctgctgctgctgctgctgct	760							
Db	109	GTCCCCAGCGATGTCTCCACCACCGCTGCTGCAACCCCTGCTGCTGCTGCTGCTGCT	168							
Qy	761	gaatgtgagccttcgcgggccaacactgatccgcacatccctcttcatcgagtccaacctg	820							
Db	169	GAATGTGGAGCCTTCCGGGGCCACACTGATCCGCATCCCTCTTATCGAGTCCAACTGG	228							
Qy	821	acgcaggaacctgaacctactgagggatggagagaaaccagcagagctccccaaagtgg	880							
Db	229	ACGCAGGATCCTGAACCTACTGAGGGGATGGAGAGAACCCAGCAGAGCTCCCAAGTTGG	288							
Qy	881	ggcccatccctggggacaagcccatcttcgtacctctctcgaactacaaggatgtga	940							
Db	289	GGCCCCATCCCTGGGACAAAGCCCATCTTCGTACCTCTCTCGAACTACAGGGATGTGCA	348							
Qy	941	gtattttggggaattgggctgggaacgctccacaaaacttcaacttgcctttgacac	1000							
Db	349	GTATTTTGGGAAATTTGGGCTGGGAAGCCCTCCACAAAATTCACCTGTTCCTTTGACAC	408							
Qy	1001	tggctctccaatctctgggtccgtccaggaagatgccacttcttcagtgctgctgctg	1060							
Db	409	TGGCTCTCCAAATCTCTGGGTCCCGTCCAGGAGATGCCATCTTTCAGTGTGCCCTGTG	468							
Qy	1061	gttacacacacgatttgatcccaaacgctctagctctctccaggccaatgggaccaagt	1120							
Db	469	GTTACACCACCGATTGATCCCAAGCCTCTAGCTCTTCCAGGCCAATGGGACCAAGTT	528							
Qy	1121	tgcattcaataaggaaactgggctgggtagatggaaactctgagcagagagaaagtgtactat	1180							
Db	529	TGCCATTCAATATGGAACTGGGGGGTAGATGGAATCTCTGAGCGAGACAAGCTGACTAT	588							
Qy	1181	tggtgaatacaagggtgcatcagtgatttcggggaggtctctctggagcccagcctggt	1240							
Db	589	TGGTGGAAATCAAGGTGCATCAGTGATTTTCGGGGAGGCTCTGTGGAGCCAGCCTGGT	648							
Qy	1241	cttcgctttggccattttgatgggatatgggctcggttttcccaattctgtctgtgga	1300							
Db	649	CTTCGTTTTTGGCCATTTTGATGGGATATTGGGCTCGGTTTCCCATTTCTGTCTGTGA	708							
Qy	1301	aggagttcgccccagtgatgactggtggagcaggggctattggaagcctgctt	1360							
Db	709	AGGAGTTTCGGCCCCCGGATGGATGTACTGGTGGAGAGGGGCTATTGGATAAGCCTGCTT	768							
Qy	1361	ctccttttactcaacagggacccctgaagcctgtagagagagctggttcctgggggg	1420							
Db	769	CTCCTTTTACTCAACAGGACCCCTGAAGCCTGTATGGAGAGAGCTGGTCTCTGGGGGG	828							
Qy	1421	ctcggaccggacactacatcccccccctcaactctgtgcccagtcacggtccctgacta	1480							
Db	829	CTCGGACCGGCACACTACATATCCACCCCTCACCTTCGTGCCAGTACGGTCCCTGCCCTA	888							
Qy	1481	ctggcagatccacatgagcgtgtgaaggtgggcccagggctgactctctgtgccaaagg	1540							
Db	889	CTGGCAGATCCACATGAGCGGTGTGAAGGTGGGGCCAGGGCTGACTCTCTGTGCCAAGGG	948							
Qy	1541	ctgtgctgcaatcctggtacacgggcaogtccctcatcacaggaccactgaggagatccg	1600							
Db	949	CTGTGCTGCCATCTGGATACGGGCACGTCCTCATCACAGGACCCACTGAGGAGATCCG	1008							
Qy	1601	ggccctgcagcagccattgggggaatcccccttctgctggctgggagtagatcatcctgtg	1660							
Db	1009	GGCCCTGCATGCAGCCATTTGGGGGAATCCCOCTTGTCTGCTGGGAGTACATCATCTCTGTG	1068							

Qy	1661	ctcggaaatcccaagctccccgcagtccttctcttcttggtggtttaacct	1720
Db	1069	CTCGAAATCCAAAGCTCCCCGAGTCTCCTTCTTTGGGGGGGTCTGGTTAACCT	1128
Qy	1721	cacggcccatgattacgtcatccaagactactcgaaaaggcgtcgcgtcttgtctgcgg	1780
Db	1129	CACGGCCCATGATTACGTCAFCACAGACTACFCGAAAATGGCGTCGCCTCTGCTTGTCGG	1188
Qy	1781	tttccaggccctggatgtccctccgcctgcagggcccttcttgatactcgtgtagctctt	1840
Db	1189	TTTCCAGGCCCTGATGTCCTCCGCCCTGCAGGCCCCTCTCGATCTCTCGTGACGTCTT	1248
Qy	1841	cttggggacgatgtggcgtctttccaccgcggggacatgaagagcagcgcgggtggg	1900
Db	1249	CTTGGGACGATGTGGCCGCTTTCGACCGGGGGACATGAAGAAGACAGCGCCGGTGGG	1308
Qy	1901	cctggcgcgctcgcactcggagcggaacctcggatggggaagagactcgcaaggcgca	1960
Db	1309	CCTGGCGCGCTCGCAC'TCGCGAGCGGAGCTCGGATGGCGAGAGACTCGCGCAGGCGCA	1368
Qy	1961	gttcccgggtgcgcgcccagaagtgaagcgcatgcagcggtggtcgcggaggtcctgct	2020
Db	1369	GTTCCCGGGTGACGCCCAAGTGAAGCGCATGCCAGCGGGTGTGCGGTAGTCTTGCT	1428
Qy	2021	accagtaaaaaatccaactatttccattgaaaaaiaaaaaa	2060
Db	1429	ACCAGTAATAAATCCACTATTTCCTATTGAGAAAAACA	1468
RESULT 4			
E3352	LOCUS	E3352	1648 bp DNA PAT 07-FEB-2001
DEFINITION	ASP5.		
ACCESSION	E3352		
VERSION	E3352_1	GI:13026985	
KEYWORDS	JP 1999155583-A/1.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	David,J.P.J.J. and Jeffrey,H.T.S.S.		
JOURNAL	ASP5		
COMMENT	Patient: JP 1999155583-A 1 15-JUN-1999; SMITHKLINE BEECHAM CORP OS Homo sapiens (human) PN JP 1999155583-A/1 PD 15-JUN-1999 PF 21-AUG-1998 JP 1998236037 PR 21-AUG-1997 US 60/056480,08-JUL-1998 US 09/111727 PI DAVID J POWERU JOHN KEI,JEFFREY HILL,TUDY SMITH PC C12N15/09,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61K48/00,PC C07K16/18, PC C12N5/10,C12P21/02,G01N33/53,G01N33/566,C12N15/00,A61K37/02, CC C12N5/00 Key Location/Qualifiers FH 1..1648 FT source /organism='Homo sapiens (human)'. FT Location/Qualifiers 1..1648 /organism='Homo sapiens' /db_xref='taxon:9606'		
BASE COUNT	343 a	492 c	459 g 354 t
ORIGIN			
Query Match	65.6%	Score 1350.4;	DB 10; Length 1648;
Best Local Similarity	99.6%;	Pred. No. 0;	
Matches 1354;	Conservative	0; Mismatches	6; Indels 0; Gaps 0;
Qy	701	gtccccagcgatgtctccaccacgcgtgctgcaacccctgctgctgctgctgctgct	760





Db 1141 TATGTGGCGCTTTCACCGCGGACATGAAGACAGCGCCCGGTGGCCCTGGCGCGC 1200  
Qy 1911 gctcgcaactcgcggagcagacctcgatgaggagagactgcgagcagctcccccggg 1970  
Db 1201 GCTCGCACTCGCGGAGCGGACCTCGGATGGGAGAGACTGCGCAGCGCGCAGTTCGCCGGG 1260  
Qy 1971 tgaaccccaagtgaagcgcacatgcgcagcgggtgggtcgcggaggtcctgctacccagtaaa 2030  
Db 1261 TGACGCCCAAGTGAAGCGCATGCGCAGCGGGTGGTTCGGGAGGTCTCTGTACCCAGTAAA 1320  
Qy 2031 aatccactatttccattgaaaaa 2060  
Db 1321 AATCCACTATTTCACATTGAGAAAAACAAA 1350

RESULT 6  
AR016588 AR016588 1299 bp DNA PAT 05-DEC-1998  
LOCUS Sequence 4 from patent US 5776759.  
DEFINITION AR016588  
ACCESSION AR016588  
VERSION AR016588.1 GI:3972865  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1299)  
AUTHORS Bandman,O. and Coleman,R.  
TITLE Two novel human cathepsin proteins  
JOURNAL Patent: US 5776759-A 4 07-JUL-1998;  
FEATURES Location/Qualifiers  
1. 1299  
/organism="unknown"  
BASE\_COUNT 258 a 385 c 373 g 283 t  
ORIGIN

Query Match 58.2%; Score 1199.2; DB 9; Length 1299;  
Best Local Similarity 94.3%; Pred. No. 0;  
Matches 1286; Conservative 0; Mismatches 3; Indels 75; Gaps 1;

Qy 697 acgagctccccagcgatgtctccacacacgcgtgctgcaacccctgctgctgctgcctc 756  
Db 3 ACGGCTCGCGAGCGATGTCTCCACCACCGCTGCTGCAACCCCGTGTGCTGCTGCCTC 62  
Qy 757 tctgaatgtgagcttccggggccacactgctccgcatccctcttcagagtcacac 816  
Db 63 TGCTGAATGTGGAGGCTTTCGGGGGCCACACTGATCCGATCCCTCTTCATCGAGTCCAAC 122  
Qy 817 ctggagcagagacctgaacctactgaggggatggagagaacacagcagagctccccaagt 876  
Db 123 CTGGAGCGAGGATCTGAACCTACTGAGGGGATGGAGAGAACACAGAGCTCCCAAGT 182  
Qy 877 tggggggccccatccccctgggagacagcccatcttcgtacctctctcgaactacagggatg 936  
Db 183 TGGGGGCCCCATCCCTCGGGACAAAGCCCATCTTCGTACTCTCTCGAACTACAGGGATG 242  
Qy 937 tgcagattttgggaaattgggtgggaacgcctcccaaaacttcactgtgctttg 996  
Db 243 TCAGTATTTTGGGGAATTTGGGCTGGGAACGCCCTCCACAAACTTCACATTGTTCCTTTG 302  
Qy 997 acactggctctcccaatctctgggtcccgctccagagagatgccaactcttcagtgccct 1056  
Db 303 ACACGTGGCTCTCCAAATCTGGGTCCCGTCCAGGAGATGCCACTCTCTCAGTGTGCCCT 362  
Qy 1057 gctggttacaccacagcttgatcccaagcctctagctcttcagagccaatgggaacca 1116  
Db 363 GCTGTTTACACCACCGATTGATCCAAAGCCTCTAGCTCTTCCAGGCCAATGGGACCA 422  
Qy 1117 agtttgcaattcaatatgaaactggcggttagatgggaatcctgagcagaggaagaagctga 1176  
Db 423 AGTTTGCCATTCAATGAACTGGCGGGTAGATGGAAATCCTGAGCGAGGACAAGCTGA 482

Qy 1177 ctattgtgggaatcaaggggtgcatcagtgattttcggggaggtctctcgggagccagcc 1236  
Db 483 CTATTTGGTGGATCAAGGGTGCATCAGTCAATTTTCGGGAGGCTCTCTGGGAGCCAGCC 542  
Qy 1237 tggcttcgcttttcccatttggatgggatatattgggctcggttttcccattctgtctg 1296  
Db 543 TGGTCTTCGCTTTTCCCATTTTGGATATTTGGGCTCGGTTTCCCATTTCTGTCTG 602  
Qy 1297 tggaaagaggttcggcccccgatggatgactggtgagagcagggggtattgataagcctg 1356  
Db 603 TGGAAAGGAGTTTCGGCCCCCGATGGATGTACTGGTGGAGCAGGGGCTATTGGGAAGCCCTG 662  
Qy 1357 tcttctccttttaacctcaacaggggacctgaagagcctgtagagagagctggctcctgg 1416  
Db 663 TCTTCTCTCTTTTACCTCAACAGGGACCTGAAGAGCCTGATGGAGGAGAGCTGGTCTCTGG 722  
Qy 1417 ggggctcggaacccggcacactacatcccaacccctcaacttcgtgcagtcacggtccctg 1476  
Db 723 GGGGCTCGGACCCCGCACACTACATCCACCCTCACCCTTCGTCGACGTACGGTCCCGC 782  
Qy 1477 cctactggcagatccacatggagcgtgtgaaggtggccagggctgactctctgtgcca 1536  
Db 783 CCTACTGGCAGATCCACATGGAGCGTGTGAAGTGGGCCCGCCAGGGCTGACTCTCTGTGCCA 842  
Qy 1537 agggctgtgctgccatcctggatacgggcagctccctcatcacagagaccactgagagga 1596  
Db 843 AGGGCTGTGCTGCCATCTGGATACGGGCACGTCCCTCATCACAGGACCCACTGAGGAGA 902  
Qy 1597 tccgggcccctgcactgagccattgggggaatcccccttgcctgctgggagtagacatcc 1656  
Db 903 TCCGGGCGCTGTCATGACGCCAATTTGGGGAAATCCCTTGTGCTGGGAGTAGATCATCTCC 962  
Qy 1657 tgtgctcgaaaatcccaagctcccccagctcctctctctctggggggtcgtggttta 1716  
Db 963 TGTGCTCGAAATCCCAAGCTCCCGCCGAGTCTCTTCTTGGGGGGGTGTGGTTTA 1022  
Qy 1717 acctcacggcccatgattacgtcatccagactactcgaatggcgtccgctcctctgtgt 1776  
Db 1023 ACCTCAGGCCCATGATTACGTATCCAGACTACTTCGAAA----- 1062  
Qy 1777 ccggtttccaggccctggatgtccctccgctgcaggggccctcttgatccctcgtgtag 1836  
Db 1063 -----TGACG 1067  
Qy 1837 tottctgggagcgtatgtggcgtcttcgacgcggggacatgaagagcagcgcgcggg 1896  
Db 1068 TCTTCTTGGGAGCGTATGTGGCCGCTTCGACCGGGGACATGAAGAGCAGCGCCCGG 1127  
Qy 1897 tgggcttgccgcgcgtgcactcgcgagcgcgcgtcggaaggagagactgcgcag 1956  
Db 1128 TGGGCTTGGCGCGCTGCGACTCGGGAGCGGACCTCGGATGGGAGAGACTGCGCAGG 1187  
Qy 1957 ggcagttcccccgggtgacgcccgaagtgaagcgcatacgcagcgggtggtcgcggaggtcc 2016  
Db 1188 CGCAGTTCCTCCCGGGTAGCGCCCAAGTGAAGCGCATCGCAGCGGGTGTGTCTCGGGAGTCC 1247  
Qy 2017 tgcataccagtaaaaaatccactatttccattgaaaaa 2060  
Db 1248 TGCTACCCAGTAAAAATCCACTATTTCATTGAAAAA 1291

RESULT 7  
AR078758 AR078758 1299 bp DNA PAT 31-AUG-2000  
LOCUS Sequence 4 from patent US 5965129.  
DEFINITION AR078758  
ACCESSION AR078758  
VERSION AR078758.1 GI:10005504  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1299)  
AUTHORS Bandman,O. and Coleman,R.

TITLE Two novel human cathespin proteins  
JOURNAL Patent: US 5965129-A 4 12-OCT-1999;  
FEATURES Location/Qualifiers  
1. .1299

BASE COUNT 258 a 385 c 373 g 283 t  
ORIGIN /organism="unknown"

Query Match 58.2%; Score 1199.2; DB 9; Length 1299;  
Best Local Similarity 94.3%; Pred. No. 0;  
Matches 1286; Conservative 0; Mismatches 3; Indels 75; Gaps 1;

QY 697 aagctccccagcagatgtctccaccacacgcgtctgcacacccctgctgtgctgtgctc 756  
Db 3 AAGCTCCCGACAGCATGTCTCCACCACCGCTGCTGCAACCCCTGCTGTGCTGCCTC 62  
QY 757 tttcgaattgtgagctctccggggccacactgacatccgacatccctcttcaatcgagtcacac 816  
Db 63 TACTGAATGTGGAGCTTCCGGGGCCACACTGATCCGATCCCTCTTCATCGAGTCCAAC 122  
QY 817 ctggcagcaggacctgaacctactgagggatgagagaccacagcagactccccaaagt 876  
Db 123 CTGGACGAGGATCTGAACCTACTGAGGGATGGAGAGACACAGCAGACTCCCAAGT 182  
QY 877 tggggccccatccccgggacaaagccatcttctgtacctctctcgaactacaggatg 936  
Db 183 TGGGGCCCCATCCCTCGGGACAGCCCATCTCGTACCTCTCTCGAATACAGGGATG 242  
QY 937 ttcagattttggggaattggctgggaacgcctcccaaaacttcaacttgcctttg 996  
Db 243 TGCACATTTTGGGGAATTTGGCTGGGAACGCCCTCCCAAAATTCCTACTGTGCTTTG 302  
QY 997 aactggctctccaatctctgggtcccgctccaggagatgccactcttccagtgtccct 1056  
Db 303 AACTGGCTCTCCAAATCTGGGTCCCGTCCAGGAGATGCCACTTCTTCAGTGTGCCT 362  
QY 1057 gctgttacacacagcattgaatccaaagcctctagctctctccaggccaatgggacca 1116  
Db 363 GCTGTTACACACCGATTGATCCAAAGCCTCTAGTCTCTCCAGGCCAATGGGACCA 422  
QY 1117 agttgccattcaatagaaactggcgggtagatggaatcctgagcagagacaagctga 1176  
Db 423 AATTGCCATCAATATGAACTGGGGGTAGATGAATCCTGAGCAGACAACTGA 482  
QY 1177 ctattgtggaatacaagggtcactagtgatttcgggagagctctctgggagccccgcc 1236  
Db 483 CTATTGGTGAATCAAGGGTGCATCAGTGATTTTCGGGGAGGCTCTCTGGAGGCCAGCC 542  
QY 1237 tggcttcgctttgcccatttgatggatattggcctcggttttcccattctgtc 1296  
Db 543 TGGCTTCGCTTTTGGCCATTTTGATGGATATTGGGCCCTCGGTTTCCCATCTGTCTG 602  
QY 1297 tgaagagagttgcccccgatgactggtgggagcaggggctattggataagcctg 1356  
Db 603 TGAAGGAGTTTCGGCCCGCGATGATGCTGTTGGAGCAGGGGCTATTGATAAGCCTG 662  
QY 1357 tcttctcttttaactcaacaggagacctgaagacactgagagagctgagggagctgtc 1416  
Db 663 TCCTCTCTTTACTCTCAACAGGAGCCCTGAAGGCTGATGGAGGAGAGCTGTGCTCTGG 722  
QY 1417 gggctcgagccggcacactacaccccccctcacctctcgtgccaagtcacagtcctc 1476  
Db 723 GGGCTCGGACCGGCACACTACATCCACCCCTACCTTCGTGCGCAGTACAGGTCCCCG 782  
QY 1477 cctactggcagatccacatggagctggaagtgggccagggctgactctctgtgcca 1536  
Db 783 CQTACTGGCAGATCCACATGGAGGCTGTAAGGTGGGCCCCAGGGCTGACTCTCTGTGCCA 842  
QY 1537 agggctgtctgcatctgatacaggacagcctccctcatcacaggaccacactgagaga 1596  
Db 843 AGSGCTGTGTCCTCTGGATACGGGCACGTCCTCTCATCAGAGGACCCACTGAGGAGA 902

QY 1597 tccgggcctcgtcagcagccatttggggaaatccctctgctgctggaggatcacatccc 1656  
Db 903 TCCGGCCCTGCATGCAGCAATTTGGGGAATCCCTTGTCTGGCTGGGAGTACATCAATCC 962  
QY 1657 tgtctcggaatacccaaaagctccccgcagctctcttcttgggggggctctggttta 1716  
Db 963 TGTGCTCGAAATCCCAAAGCTCCCGCAGTCTCTTCTCTTGGGGGGTCTGGTTTA 1022  
QY 1717 acctcacggccatgattacgtcatccagactactgaaatggcgtccgcctctgcttgt 1776  
Db 1023 ACCTACAGGCCCATGATTACGTATCCAGACTACTCGAAA----- 1062  
QY 1777 ccggtttccaggccctggatgtccctccgcctgcagggcccttctggtatcctcgtgacg 1836  
Db 1063 -----TGACG 1067  
QY 1837 tctcttggggaactatgtggccgtctctcgaccgctgggacatgaagacagcagcccg 1896  
Db 1068 TCTTCTTTGGGACGATATGTGCCCTCTTCGACCCTGGGACATGAAGAGCAGCGCCCGG 1127  
QY 1897 tggccttggcgcgcgtcgcactcgcagcagcagcactcgatgggagagactgcgcagg 1956  
Db 1128 TGGCCCTTGGCGCGCTCGCACCTCGCGAGCGGACCTCGGATGGGGAGAGACTGCGCAGG 1187  
QY 1957 cgcagttcccccgggtgacgcccgaagtgaagcgcagcgcagcgggtggtcgcgaggtcc 2016  
Db 1188 CGCAGTTCCCGGGTGACGCCCAAGTGAAGCGATGCGCAGCGGGTGTGCGGAGAGTCC 1247  
QY 2017 tgtaccacgataaaatccactatttccattgaaaaaataaaaaa 2060  
Db 1248 TGTACCCAGTAAATCCACTATTTCATTGAAAAAATAAAAAA 1291

RESULT 8

A70198 LOCUS A70198 1375 bp DNA PAT 07-MAY-1999  
DEFINITION Sequence 2 from Patent WO9811236.  
ACCESSION A70198  
VERSION A70198.1 GI:4774612  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1375)  
AUTHORS Hill,J., Kay,J. and Powell,D.  
TITLE ASPARTIC PROTEASE  
JOURNAL Patent: WO 9811236-A 2 19-MAR-1998;  
HILL JEFFREY (GB)  
FEATURES Location/Qualifiers  
source 1. .1375  
/organism="unidentified"  
/db\_xref="taxon:3264"  
BASE COUNT 255 a 419 c 396 g 305 t  
ORIGIN

Query Match 55.9%; Score 1151.8; DB 9; Length 1375;  
Best Local Similarity 90.8%; Pred. No. 1.3e-290;  
Matches 1238; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 698 cgcgtccccagagatgtctccaccacgcgtgctgaacccctgctgctgctcctet 757  
Db 13 CGGCTCCCCAGCGATGCTCCACACCGTGTCTGCTACCTTTGCTGCTGCTGCT 72  
QY 758 gctgaattggagcctccggggccacactgacccatccctcttcacgagtcacac 817  
Db 73 GCTGATGTGGAGCTGCTGGGCCACACTGATCCGGATCCCTCTTCGTCAAGTCCACCC 132  
QY 818 tgaacagcagccctgaacctactgagggatggagagaccacagcagactcccaagtt 877  
Db 133 TGCACGACGAGCCCTGAACCTACTGAGGGGATGGGGAAAAACGACAGAGTCCCCAAGTT 192  
QY 878 gggggcccatccccctggggaacaagcccatcttcgtacctctctcgaaactacaggtatg 937









```

Qy 1491 cacatggagcgtgtaaggtggccagggtgactctctgtcgaagggtgtgtgctgcc 1550
Db 781 CACATGGAGCGTGTGAAGGTGGCTCACGGCTGACTCTGTGCCCAGGGCTGTGCTGCC 840
Qy 1551 atcctggatacgggcaagctccctcatcacagagaccactgaggagatccgggcccctgcat 1610
Db 841 ATCTGTGGATACAGGCACACCTGTATCTGTAGGACCCACTGAGGAGATCCGGGCCCTGTGAT 900
Qy 1611 qcagccattgggggaatccccctgtgctggtgggaggtacatcctctgtctcggaatc 1670
Db 901 CGAGCCATTGGGGGAATCCCCCTGCTGGCTGGGAGTACATCATCCGGTGTCTAGAAATC 960
Qy 1671 ccaaagctccccgcagtcctctctctctctctctctctctctctctctctctctctctct 1730
Db 961 CCAAAGCTCCCCGCGAGTCTCACTCTCTATTTGGGGGGTCTGTTTAATCTCACGGCCAG 1020
Qy 1731 gattacgtcatccagactaactcgaataggcgctcgccctctctctctctctctctctct 1790
Db 1021 GATTACGTCATCCAGTGTGCTCAGGGTGACGTCCGCCCTCTCTGTCGGGTCTCCGGGCC 1080
Qy 1791 ctggatgtccccctcgctgcagggccctctctggtatccctcgctgacgtctctctctctct 1850
Db 1081 TTGGACATCGCTTCCCTCTCAGTACCTGTGTGGATCTTCGGCGACGTTTCTTGGGGCG 1140
Qy 1851 tatgtggccgtctctgacccgvggacatgaagagcagcccggtggtggtggtggtggtggt 1910
Db 1141 TATGTGACCGTCTTCGACCGCGGACATGAAGAGCGCGCAGGAGTGGGACTGGCGCGC 1200
Qy 1911 gctcgcaactcggagcgggaactcggatgggagagagactgcgagggcagctctcccggg 1970
Db 1201 GCTCGCCCTCGCGAGCGGACCTGGGAAGCGCGAGAGCGCGCAGGCGCATACCGCGGG 1260
Qy 1971 tqagcccaagtgaagcgcgtgcagcgggtggtcgcgagggtcctgctaccagtaaa 2030
Db 1261 TGCGCCCGCAGTGATGCGCATGCGCACCG---GGTAGCGGAGCTAGCGCTACTCAGTAAA 1317
Qy 2031 aatccactatttccattgaaaaaataaaaaa 2060
Db 1318 AATCCAAATATTTCATTGAAAAAATAAAAAA 1347

RESULT 13
A70212
LOCUS A70212 1347 bp DNA PAT 07-MAY-1999
DEFINITION Sequence 16 from Patent WO9811236.
ACCESSION A70212
VERSION A70212.1 GI:4774625
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1347)
AUTHORS Hill,J., Kay,J. and Powell,D.
TITLE ASPARTIC PROTEASE
JOURNAL Patent: WO 9811236-A 16 19-MAR-1998;
HILL JEFFREY (GB)
FEATURES
source
1..1347
/organism="unidentified"
/db_xref="taxon:3264"
BASE COUNT 262 a 386 c 388 g 306 t 5 others
ORIGIN

```

Query Match 46.18; Score 949.2; DB 9; Length 1347;  
 Best Local Similarity 89.4%; Pred. No. 1.5e-237;  
 Matches 1091; Conservative 0; Mismatches 118; Indels 11; Gaps 6;

```

Qy 698 cgcgtccccagcagtgatctccaccacgcgtgctggaacccct-gctgctgtctgctgcctc 756
Db 44 CGGTCCCGCCAGCATGTCTCCACCACACTGCTGCTGTACTACCTTCAGTCTGTGTCGCTC 103
Qy 757 tgcgtgaatgtgagccttcggggggccacactgatccgcgcatccctctctctatcagagtc 816

```

```

Db 104 TCCTGAATGTGGAGCGTCTGCTGGGGCCACACTGATCCGATCCCTCTTCTGTCGAAGTCCACC 163
Qy 817 ctgagcgagagaccctgaacctactagagggatggagagaaccagagagctccccaagt 876
Db 164 CTGGACGAGACACCTGAACCTACTAGAGGGATGGGAAAAACCAAGAGAGCTCCCCAAGT 223
Qy 877 tggggccccctccccctggggaagcccatcttcgtacctctctcgaactacagggatg 936
Db 224 TGGGGCCCCCATCCCTGCGGACAAAGCCTGCTCGGTACCTCTCTCCAAATTCCTTGGATG 283
Qy 937 tgcagtatatttggggaaattgggctgggaacgcctcccaaaacttccactgttgccttg 996
Db 284 CCCAGTATTTTGGGAAATTTGGCTGGGAACGCCCTCCACAAAACCTCACTGTGTGCTTTG 343
Qy 997 aactggctcctcaactctctgggtcccgctccagagatgcaactcttccagtgtgcct 1056
Db 344 ACATGGGTCTCTCCAACTCTGGGTCCCGTCCAGAGATGCCACTTCTTCACTGTGCGCT 403
Qy 1057 gctggttacacacacgatttgatcccaagcctctagctccttccagcccaatggagcca 1116
Db 404 GCTGGTTCACACACCGCTTCAATCCCAATGCCCTCCAGCTCCTTCAAGCCCACTGGGACCA 463
Qy 1117 agttgcatatcaatatggaactggcggtggtagatggaaatcctgagcgagacagaagtga 1176
Db 464 AGTTTGCCATTCAAGTATGGAATGGCGGGTAGATGGAATCCTGAGTGAGGACAAGCTGA 523
Qy 1177 statgtgggaatcaagggtgcatcagtgattttcggggagcctctctcggagaccagcc 1236
Db 524 CTATTGGTGGAAATCAAGGGTGCATCCGTGATTTTCGGGGAAGCTCTGTGGGAATCCAGCC 583
Qy 1237 tggctctcgtcttgcctcatttgatgggatattggcgtcggttttcccatctgtctg 1296
Db 584 TGGTCTTCACTGTTCCTCCCGCCGATGGGATATTTGGGCTCGGTTTTCCCATTCGTCTG 643
Qy 1297 tggagagggttcggcccccgatggatgtactggtgagcagggggtatgggaagcctg 1356
Db 644 TGGAAAGAGTTTCGGCCCCCGCTGGATGTACTGTTGAGCAAGGGCTATTGGATAAGCGCTG 703
Qy 1357 tctctcctcttaccctcaacaggaccctgaagacctgaagcctgagagagctggcctgg 1416
Db 704 TCTTCTCTCTTTTATTTCACAGGGACCTGTAAGTGGTTAATGGAGGAGAGCTGTGCTTGG 763
Qy 1417 ggggctcgagcccgacactacatcccaacccctcactctgtgccagtcacgctcctg 1476
Db 764 GGGCTCGGACCCGCGCACACTACATCCACCCCTCAACTCTGTCGCCAGTACGGTCCCG 823
Qy 1477 cctactggcagatccacatggagcgtgtgaaggt-gggccaggggctgactctctgtgcc 1535
Db 824 CCTACTGGCAGATCCACATGGAGCGTGTGAAGGTGGGGCCCAAGGCTGACTCTCTGTGCC 883
Qy 1536 aaggcgtgtgctgccatcctggatacgggcagctccc---tcatcacaggaccactgag 1592
Db 884 AAGGG-TGTGTGCCATCTGGATACAGCAGTACCTACCTGGTATCACAGGACCCCATGAG 942
Qy 1593 gagatccgggcccctgcatgagccattggggaaatccccctgtgctgctggggagtagatc 1652
Db 943 GAGATCCGGGCCCTGCATGCAGCCATTGGGGAAATCCCCCTTGTGCTGGGAGTACATC 1002
Qy 1653 atcctgtgtcggaaatcccaagctcccccgagctcctctctcttcttgggggggctctgg 1712
Db 1003 ATCCGGTGTCTAGAAATCCCAAGCTCCCGCAGTCTCACTCTCATTTGGGGGGGTCTGG 1062
Qy 1713 tttaacctcaaggcccatgattacatccagactactcgaata---tggcgctcgcctc 1769
Db 1063 TTTAATCTCACGGGCCAGGATTTACCTCATCCAGACTACTCGAAAAGGGTGACGTCGCGCTC 1122
Qy 1770 tgcctgtccggtttctccagccctggtgctcctcgcctgcagggccctcttggtatcctc 1829
Db 1123 TGCTTGTCCGGCTTCAGGGCCCTTGGACATCGCTCGGGCTGAAGGACCTGTCTGGATCTC 1182
Qy 1830 ggtgagctcttcttggggagcgtatgtggcctctcttcgaccgcggggagcatgaagagcagc 1889

```

Db 1183 GCGAGATTTTGGGAA--TATGTGACCGTTTGTACCGGGGACATGAAGAGCGGCGA 1240  
Qy 1890 gcccgggtggccctgagcg 1909.  
Db 1241 ACGGAGTTGACTTGGGG 1260

RESULT 14  
AX046346 1360 bp DNA PAT 24-NOV-2000  
LOCUS Sequence 13 from Patent WO0011168.  
DEFINITION AX046346  
ACCESSION AX046346  
VERSION AX046346.1 GI:11344334  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1360)  
AUTHORS Lemischka, I. and Moore, K.  
TITLE Genes that regulate hematopoietic blood forming stem cells and uses thereof  
JOURNAL Patent: WO 0011168-A 13 02-MAR-2000;  
Princeton University (US)  
FEATURES  
Source 1..1360  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
BASE COUNT 293 a 385 c 355 g 327 t  
ORIGIN

Query Match 38.7%; Score 797.2; DB 9; Length 1360;  
Best-Local Similarity 75.6%; Pred. No. 9.4e-198;  
Matches 1034; Conservative 0; Mismatches 318; Indels 16; Gaps 3;

Qy 694 agaacgctccacagcagctgtctccacacgcctgtgcgaacccctctctctgctgtgc 793  
Db 1 AGCTAGTCCCGAGAGTGGCCACTACTGTG-----CTGCTGCTGGCGTCG 51

Qy 754 ctctgtgaatgtggagccttcggggccacactgatccgcactctcttcacagtc 813  
Db 52 TGTGGGGAATTTGGAGCCTCAGAGGCCAACTGATGCCGTCTCCCTTCAACGAATCC 111

Qy 814 aactgacacagacacctgaactactagggatgagagaaacacagacagactcccca 873  
Db 112 ACTTGGACACAGAACTTAAACCCACTGAATGGATGGGAAACAGCTGGCAGAGCTTTCTA 171

Qy 874 agttggggcccatccctcctgggggacaaagcccatctctctacactctctcgaaactacagg 933  
Db 172 -----GGACCTCCACCTCTGGTGGCAACCCCTCTTGTGGCTCTCTCCAGTTCAATGA 225

Qy 934 atjtgagtagtttgggaaattggctggggaaacgcctccacaaaacttcactgttgct 993  
Db 226 ACACCCAGTATTTGGAACTATTTGGTTTGGGAACGGCTCTCTAGAAATTTACCCGTTGTCT 285

Qy 994 ttjacactggctctccaactctgggtcccgctccggagatcccaactcttcagtgctc 1053  
Db 286 TTACACGGGTCTTCCAACTTGTGGTCCGTCACAGAGATGTCATTCTTCAGTTGG 345

Qy 1054 ccgctggttacaccacagatttgatcccaaaagcctctagctctccaggccaatggga 1113  
Db 346 CATGCTGGTTTACCATCGCTTTAATCCCAAGGCCTCCAGCTCCTTCAGGCCCAATGGGA 405

Qy 1114 ccagtttgcattcaatgaactggcggttagatgaactcctgagcaggacaagc 1173  
Db 406 CCAAGTTGCCATTAGTAGTGGACCGCGCTGAGCGGAATPCCGTGACGAGGAATTC 465

Qy 1174 tgactattggtgaatacaagggtgcacatcagtgattttcggggaggtctctgaggacca 1233  
Db 466 TGACTATCGGGGATCCACGATGCTTTTGTGACATTGGAGAGGCTCTGTGGAGCCCA 525

Qy 1234 gcttggtctctgttttgcccatattgaggatattgggctcggctcgggtttcccatctgt 1293

Db 526 GCCTGATCTTGTCTTAGCCACATTTGATGGGATCTTGGGCTCGGCTTCCCCTACTGG 585  
Qy 1294 ctgtggaagagtgctggcccccgatggtactgtgtggagcagggctatttgataagc 1353  
Db 586 CTGTGGGGGAGTTTCAGCCTCCGCTGGATGGATGGTGGAGCAAGGGCTGCTGCGAGAAC 645  
Qy 1354 ctgtctctctcttttactcaacagggacacctgaagagcctgatggagagagctgtgc 1413  
Db 646 CCGTCTTCTCTTCTTACCTCAACAGGATTTCTGAAGGGTCTGATGGGGGAGAGCTGTGC 705  
Qy 1414 tggggggtcggaccggcgcacactacatccaccctccactctcgttcgctcagtcacgctcc 1473  
Db 706 TAGGGGCTCAGACCCCGCTCACTACGTACTCCTCCCTCACTTATACAGTACCAATCC 765  
Qy 1474 ctgcctactggagatccacatggagcgtgtgaagtgggccacagggctgactctgtg 1533  
Db 766 CTGCCTACTGGCAGGTCACATGAGAGTGTGAAGGTGCGCAGAGGCTGAGGCTGTGTG 825  
Qy 1534 ccaagggtcgtgctgccactcctgtgacggcgacgtccctcatcacagagccactgagg 1593  
Db 826 CCCAGGGCTCAGTGCATCTTAGACACAGGCACATCCCTTCTCTGAATGGCAGTACTTCA 885  
Qy 1594 agatccgggcccgtgcacgcccattgggggaatcccccttgcctgtgggtgggagctatcatca 1653  
Db 886 AGATCCGGGCTTGAATAAGCCATTGGGGGATATCCCTTCTCTGAATGGCAGTACTTCA 945  
Qy 1654 tctgtgctcggaatcccaagctcccgacgtctctctctctctctctctctctctctctctgtg 1713  
Db 946 TTCAGTCTTCCAAAGACGCCAACGCTTCCCGCTGTCTCTTCCACCTTGGTGGAGTCTGTT 1005  
Qy 1714 ttaacctcacggcccatgattacgtcatccagactactcgaaatggctcgcctctgct 1773  
Db 1006 TTAACCTCACAGGCCAGGACTATGTCATCAAGATTCTCAGACGATGTGGGCTCTGCC 1065  
Qy 1774 tgcctgggttccaggccctggatgtctccctccgcctcagggccctctctctctctctctgtg 1833  
Db 1066 TGTGGGCTTCCAAAGCCTTGGATATCCCAAGCCTCGGGACCCCTCTGATCTGATCTTGGGG 1125  
Qy 1834 acgtctcttggggagcgtatgtggcgtctctgacgcgggagacatgaagacagccccc 1893  
Db 1126 ACCTCTTTTGGGCCCCCTATGTGGCTGTCTTTGACCGCTGGGACAAAACAGTGGCCCCG 1185  
Qy 1894 ggggtggcctggcgcgcctcgcactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1953  
Db 1186 GCGTGGAGTGGCGCTGCTCAGTCTCGTTCAACAGACCGGGCAGAAAAGAGTACGC 1245  
Qy 1954 aggcgcagttcccccgggtgaccccccaagtgga-aggcgcagtcgcagcgggtgtgcggag 2012  
Db 1246 AGCGCAGTCTTCAAAAGACGCCCTTGGTTAGGGTTACAAGCTCACCGGCCACAGCAGCT 1305  
Qy 2013 gtctgtctaccagtaaaaaatccactatttccattgaaaaaataaaaaa 2060  
Db 1306 ATGCTTCTTCCAATTAACAAACTAAATAAAAAAATAAAAAAATAAAAAA 1353

## RESULT 15

AX046376

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Sequence 43 from Patent WO0011168.  
AX046376  
AX046376.1 GI:11344363

PAT

24-NOV-2000

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1360)

Lemischka, I. and Moore, K.

Genes that regulate hematopoietic blood forming stem cells and uses thereof

Patent: WO 0011168-A 43 02-MAR-2000;

Princeton University (US)



```
FEATURES          Location/Qualifiers
  source          1..1360
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
BASE COUNT       293 a 385 c 355 g 327 t
ORIGIN
Query Match      38.7%; Score 797.2; DB 9; Length 1360;
Best Local Similarity 75.6%; Pred. No. 9 4e-198;
Matches 1034; Conservative 0; Mismatches 318; Indels 16; Gaps 3;
Qy 694 agaagcggtccccagcagatgtctccacacacgcgtgctgcaacccctgctgctgctgctgctg 753
Db 1 AGCTCTAGTCCCCAGAGATGCGCACACTACTGCTG-----CTGCTGCTGCTGCTGCTG 51
Qy 754 ctctgctgaatgtgagacctccggggccacacatgacgcgacatccctcttcatcgagtcc 813
Db 52 TGCTGGGGAATTTGGAGCCTGAGGAGGCCAAACTGATCCGTGTCCTCTTCAACGAATCC 111
Qy 814 aacctgacgcagacacctgaacctactgagggatgagagaaaccagcagagctcccca 873
Db 112 ACCTTGGACACAGAACTTTAAACCCACTGTAATGATGGGAACAGCTGGCAGAGCTTTCTA 171
Qy 874 agttgggggccccatccccctgggacaaagcccatctctgtacctctctcgaactacaggg 933
Db 172 -----GGACCTCCACCTCTGCTGGTGGCAACCCCTCTTTGTGTCCTCTCTCCAAGTTCATGA 225
Qy 934 atgtgcagatatttggggaaattgggctgggaacgcctcccaaaactctactgttgcct 993
Db 226 ACACCCAGTATTTTGGAACTATTGGTTTGGGAAGCCCTCCCTCAGAAATTTACCCGTTGTCT 285
Qy 994 ttgacctggtccctccaaactctgggtccggtccagagagatgccactctctcagtgtgc 1053
Db 286 TTGACACGGGTCTTCACACTGTGGGTTCGCTCCACGAGATGTCATTTCTCAGTTTGG 345
Qy 1054 cctgctggttacacacacgatttgatcccaagcctctagctctctccagggcccaatggga 1113
Db 346 CATGCTGGTTTTCACCATCGCTTAATCCAAAGGCTCCAGCTCCTTTCAGGCCCAATGGGA 405
Qy 1114 ccaagtttgccatcaatatagaactggcgggtagatggaatccctgagcagagacaagc 1173
Db 406 CCAAGTTTGGCCATTACGATATGGGACCGGGCGGCTGAGCGGAATCCTGAGCCAGGACAAATC 465
Qy 1174 tgactattggtggaatcaagggtcatcagtgatttccggggaggtcctctgggagccca 1233
Db 466 TGACTATCGGGGGGATCCACCATGCTTTTGTGACATTTGGAGAGGCTCTGTGGAGCCCA 525
Qy 1234 gcctggtcttcgttttggccattttgatgggatatgggacctcggttttcccatctgt 1293
Db 526 GCCTGATCTTTTGTCTTAGCCCATTTGATGGATCCTGGGCTCGGCTTCCCACTCTGG 585
Qy 1294 ctgtggaaggagttcgcccccgatggtgactggtggaagcaggggctatggataaagc 1353
Db 586 CTGTGGCGGAGTTTCAGCCTCCGCTGGATGGGTGGTGGAGCAAGGGCTGTGGAGAAAC 645
Qy 1354 ctgtctctctcttttacctcaacaggacctgaagacctgatggagagagctggtgcc 1413
Db 646 CCGTCTTCCTTTTACCTCAACAGGGATTTCTGAAGGCTCTGATGGGGGAGAGCTGGTCC 705
Qy 1414 tggggggctcgagccggcacactacatcccaacccctcacctctggtgccagtcacggtcc 1473
Db 706 TAGGGGGCTCAGACCCGCTCCTACGTAGTACCTCCCTCACTCCATCCAGTCACCATCC 765
Qy 1474 ctgctactgagagatccacatgagcgtgtgaagggtggcccgagggctgactctctgtg 1533
Db 766 CTGCCCTACTGGCAGGTCCACATGGAGAGTGTGAGGTCTGGACACAGGGCTGAGCCTCTGTG 825
Qy 1534 ccaagggctgtgtgcacctctgatacggcagctccctcatcacagaccactaagg 1593
Db 826 CCCAGGGCTGACGTGCCATCTTAGACACAGGCACATCCCTCATCAGGAGCTAGTGAGG 885
Qy 1594 agatccgggcccctgcatgcagccattgggggaatcccccttgggtggggagtagatca 1653
```

```
Db 886 AGATCCGGGCTTGAATAAAGCCATTGGGGATATCCCTTCTGAATGGCGAGTACTTCA 945
Qy 1654 tccctgctgcggaataatcccaaaagctcccccagctctcctctctcttgggggggtcctggt 1713
Db 946 TTCAGTGTTCCTCAAGACGCCAACGCTTCCCTGCTCTCCCTTGGTGGAGTCTGCT 1005
Qy 1714 ttaacctcaegggcccatgattacgtcatccagactactcgaataggcgtccgctcctgct 1773
Db 1006 TTAACCTCACAGGCCAGGACTATGTCATCAAGATTTCTCAGAGCGATGTTGGCCTCTGCC 1065
Qy 1774 tgcctgggtttccaggccctggatgtccctccgctgcagggccctcttgggtatccctcggtg 1833
Db 1066 TGTGGGGCTTCCAAGCCTTGGATATCCCAAGCCTGCGGACCCCTCTGATCTCTTGGGG 1125
Qy 1834 acgtctcttgggagcgtatgtggccgtcttcgaccgcggggacatgaagagcagcgccc 1893
Db 1126 ACGTCTTTTGGGGCCCTTATGTGGCTGTCTTTGACCGTGGGGACAAAAACGTCGGCCCGC 1185
Qy 1894 ggggtgggctgctggcgctgcactgcgagcgagcgacctcgatggggagagactgcgc 1953
Db 1186 CGTGGGACTTGGCGGTGCTCAGTCTCGTTCAACAGACCGGGCAGAAAGAGGACTAGCG 1245
Qy 1954 aggcgcagttccccgggggtgacgccccaaqtga-agcgcatgcgcagcggggtggtcgcgag 2012
Db 1246 AGGCGCAGTTCTTCAAAAAGACGCCCTGCTTAGGGTACAAGCTCACCGGGCCACACAGCCT 1305
Qy 2013 gtctgctaccagctaaataatccactattccatttgaaaaaataaaaaa 2060
Db 1306 ATGCTTCTTCCAAATTAACAACTAAAAAATAAAAAAAAAAAAAAAAAA 1353
```

Search completed: November 17, 2001, 14:54:41  
Job time: 7635 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:50:26 ; Search time 222.76 Seconds  
(without alignments)  
5806.596 Million cell updates/sec

Title: US-09-700-770-3  
Perfect score: 2060  
Sequence: 1 cttgagagctctcaataact.....ttccattgaaaaaataaaaaa 2060

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : → N\_Geneseq\_0601.\*
- 1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*
  - 2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*
  - 3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*
  - 4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*
  - 5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*
  - 6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*
  - 7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*
  - 8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*
  - 9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*
  - 10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*
  - 11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*
  - 12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*
  - 13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*
  - 14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*
  - 15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*
  - 16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*
  - 17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*
  - 18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*
  - 19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*
  - 20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*
  - 21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*
  - 22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	2060	100.0	2060	21	AAZ29722 Human lung specifi
2	2060	100.0	2061	20	AAZ87152 Human protease HUP
3	1356.8	65.9	1910	19	AAV27038 Human napsin B CDN
4	1261.4	61.2	1263	21	AAZ50231 Human Aspartic Pro
5	1199.2	58.2	1298	19	AAV31665 Nucleotide sequenc
6	1151.8	55.9	1375	19	AAV28623 Human aspartic pro
7	1142	55.4	1353	19	AAV27036 Human napsin A CDN
8	1122.6	54.5	1329	21	AAZ50232 Human Aspartic Pro
9	949.2	46.1	1347	19	AAV28624 Human aspartic pro
10	797.2	38.7	1360	21	AAZ94089 Haematopoietic ste
11	797.2	38.7	1360	21	AAZ94118 Haematopoietic ste

12	716.4	34.8	1645	19	AAV27037 Human napsin A gen
13	602.4	29.2	1041	21	AAV18208 Lung cancer associ
14	397.2	19.3	481	19	AAV28634 Human aspartic pro
15	387.6	18.8	407	21	AAV27038 Human napsin B CDN
16	386.6	18.8	407	20	AAV27038 Human napsin B CDN
17	360.2	17.5	474	20	AAV27038 Human napsin B CDN
18	325	15.8	408	20	AAV27038 Human napsin B CDN
19	322.6	15.7	402	19	AAV27038 Human napsin B CDN
20	316.6	15.4	1300	21	AAV27038 Human napsin B CDN
21	316.6	15.4	1353	21	AAV27038 Human napsin B CDN
22	316.6	15.4	2038	16	AAV27038 Human napsin B CDN
23	316.6	15.4	2038	19	AAV27038 Human napsin B CDN
24	316.6	15.4	2465	14	AAV27038 Human napsin B CDN
25	309.4	15.0	377	19	AAV27038 Human napsin B CDN
26	307.8	14.9	2124	20	AAV27038 Human napsin B CDN
27	304	14.8	555	19	AAV27038 Human napsin B CDN
28	300.4	14.6	402	19	AAV27038 Human napsin B CDN
29	289.2	14.0	299	20	AAV27038 Human napsin B CDN
30	289.2	14.0	299	21	AAV27038 Human napsin B CDN
31	287.8	14.0	1714	21	AAV27038 Human napsin B CDN
32	284.4	13.8	1719	21	AAV27038 Human napsin B CDN
33	264.8	12.9	385	19	AAV27038 Human napsin B CDN
34	249.2	12.1	289	21	AAV27038 Human napsin B CDN
35	247	12.0	452	19	AAV27038 Human napsin B CDN
36	222.6	10.8	1459	6	AAV27038 Human napsin B CDN
37	220.6	10.7	296	19	AAV27038 Human napsin B CDN
38	219.4	10.7	1358	18	AAV27038 Human napsin B CDN
39	216	10.5	227	21	AAV27038 Human napsin B CDN
40	215	10.4	215	21	AAV27038 Human napsin B CDN
41	198.8	9.7	808	20	AAV27038 Human napsin B CDN
42	191.2	9.3	500	19	AAV27038 Human napsin B CDN
43	186.8	9.1	215	21	AAV27038 Human napsin B CDN
44	184	8.9	186	21	AAV27038 Human napsin B CDN
45	181.4	8.8	349	19	AAV27038 Human napsin B CDN

ALIGNMENTS

RESULT	1
AAZ29722	
ID	AAZ29722 standard; DNA; 2060 BP.
AC	AAZ29722;
XX	
DT	27-MAR-2000 (first entry)
XX	
DE	Human lung specific gene Lngl05.
XX	
KW	Lung Specific Gene; LSG; Lngl05; human; diagnostic marker;
XX	
OS	Homo sapiens.
XX	
FT	Key
XX	
FT	CDS
XX	
FT	Location/Qualifiers
XX	
FT	711..1973
XX	
FT	/*tag= a
XX	
FT	/product= "LSG Lngl05 protein"
XX	
PN	WO9960160-A1.
XX	
PD	25-NOV-1999.
XX	
PF	12-MAY-1999; 99WO-US10344.
XX	
PR	21-MAY-1998; 98US-0086212.
XX	
PA	(DIAD-) DIADEXUS LLC.
XX	
PI	Yang F, Macina RA, Sun Y;
XX	
DR	WPI; 2000-116320/10.
XX	
DR	P-PSDB; AAY44457.

XX A new method for diagnosing, monitoring and staging lung cancer  
XX  
XX Claim 6; Pages 34-35; 40pp; English.  
XX  
CC The present sequence is a lung specific gene (LSG) Lngl05 from human  
CC clone ID 3107312. The LSG has high level of tissue specificity for lungs  
CC and is overexpressed in cancerous tissues. The sequence serves as a  
CC diagnostic marker for detecting, monitoring, staging and prognosticating  
CC lung cancer. The diagnosis involves comparing levels of LSG in samples  
CC obtained from patient and normal control.  
XX  
SQ Sequence 2060 BP; 458 A; 573 C; 537 G; 492 T; 0 other;  
  
Query Match 100.0%; Score 2060; DB 21; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2060; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 cttgagagctctcaataacttggtcatggtgaagccgacgaataactgaatgattt 60  
Db 1 cttgagagctctcaataacttggtcatggtgaagccgacgaataactgaatgattt 60  
  
QY 61 ttgacacagaggttgacagcctcgagatcggaacattctctctcgccaccatgac 120  
Db 61 ttgacacagaggttgacagcctcgagatcggaacattctctctcgccaccatgac 120  
  
QY 121 ctgaaggttccaaaacttcagcgagcagctctgaagaatcctgtgaatgcccgttcc 180  
Db 121 ctgaaggttccaaaacttcagcgagcagctctgaagaatcctgtgaatgcccgttcc 180  
  
QY 181 ctttaaataccagacagttgaaaaattacacgaataattatttttttcccttaaat 240  
Db 181 ctttaaataccagacagttgaaaaattacacgaataattatttttttcccttaaat 240  
  
QY 241 ctgaagatactacctggtttatttcttaaatgaattggtggaactcctttatgatt 300  
Db 241 ctgaagatactacctggtttatttcttaaatgaattggtggaactcctttatgatt 300  
  
QY 301 ctgcagacactgtaataataaccagagaaacagcttctgactcgaaattctggtctcac 360  
Db 301 ctgcagacactgtaataataaccagagaaacagcttctgactcgaaattctggtctcac 360  
  
QY 361 tccatccccctccatggacaaatgagtcagagtaagccttagatcccttaataagtt 420  
Db 361 tccatccccctccatggacaaatgagtcagagtaagccttagatcccttaataagtt 420  
  
QY 421 ttggcccaagcccggttccattctctagcaactgaactgacgttgcagccgaggtttggacat 480  
Db 421 ttggcccaagcccggttccattctctagcaactgaactgacgttgcagccgaggtttggacat 480  
  
QY 481 acctcatgtagatgtgtgtcaactttgacattcttaccattccaaagattacatcca 540  
Db 481 acctcatgtagatgtgtgtcaactttgacattcttaccattccaaagattacatcca 540  
  
QY 541 ttgagtaggttcgaacagctagagctggcgctccggaaggttacttactttgtccacaca 600  
Db 541 ttgagtaggttcgaacagctagagctggcgctccggaaggttacttactttgtccacaca 600  
  
QY 601 gtatgatgtggaactcttcagcgcatagacacttaatttgggaagaataccaggttt 660  
Db 601 gtatgatgtggaactcttcagcgcatagacacttaatttgggaagaataccaggttt 660  
  
QY 661 tcaacacagagatgatgaggttatgtatgctgacagaacgctcccgagcatctccac 720  
Db 661 tcaacacagagatgatgaggttatgtatgctgacagaacgctcccgagcatctccac 720  
  
QY 721 caccgctgtgcacccctctgctgtgctgtctgtctgtctgtgaatgtgagccttccgggg 780  
Db 721 caccgctgtgcacccctctgctgtgctgtctgtctgtctgtgaatgtgagccttccgggg 780  
  
QY 781 ccacactgatccgcatccctcttcatagatccaactggacgcaggaaccctgaactac 840  
Db 781 ccacactgatccgcatccctcttcatagatccaactggacgcaggaaccctgaactac 840

Db 781 ccacactgatccgcatccctcttcatcgagtcacaaacctggagcaggaaccctgaacctac 840  
QY 841 tgaggggagtgagagaaaccagacagagctccccaagttggggggcccatccctctggggaca 900  
Db 841 tgaggggagtgagagaaaccagacagagctccccaagttggggggcccatccctctggggaca 900  
QY 901 agcccatcttctgacctctctcgaaactacagaggtgacatattttgggaaatgggc 960  
Db 901 agcccatcttctgacctctctcgaaactacagaggtgacatattttgggaaatgggc 960  
QY 961 tgggaacgctcccaaaaacttcaactgttgccctttgacactggctccctccaactctcgg 1020  
Db 961 tgggaacgctcccaaaaacttcaactgttgccctttgacactggctccctccaactctcgg 1020  
QY 1021 tcccgctcaggagatgccacttcttcagttgcctctgctggttacacacacgattgatc 1080  
Db 1021 tcccgctcaggagatgccacttcttcagttgcctctgctggttacacacacgattgatc 1080  
QY 1081 ccaaagcctctagctccttccagcccaatgggacaaagtttgccattcaatatgaaactg 1140  
Db 1081 ccaaagcctctagctccttccagcccaatgggacaaagtttgccattcaatatgaaactg 1140  
QY 1141 ggcgggttagatggaatcctcgagcgaggaacaagtgtactattggtggaatcaaggtgcat 1200  
Db 1141 ggcgggttagatggaatcctcgagcgaggaacaagtgtactattggtggaatcaaggtgcat 1200  
QY 1201 cagtgtatttcggggaggtctcttgagagccagcctggtctctctgtttgcccattttg 1260  
Db 1201 cagtgtatttcggggaggtctcttgagagccagcctggtctctctgtttgcccattttg 1260  
QY 1261 atgggatttggtggcctcggttttcccaattctgtctgtggaagaggttcggcccccagtg 1320  
Db 1261 atgggatttggtggcctcggttttcccaattctgtctgtggaagaggttcggcccccagtg 1320  
QY 1321 atgtactgttgagcagagggcctattggaataagcctgtctctcttcttaactcaacaggg 1380  
Db 1321 atgtactgttgagcagagggcctattggaataagcctgtctctcttcttaactcaacaggg 1380  
QY 1381 acctgaagagcctgtgagggagagctggtcctgggggctcgaccggcgacactaca 1440  
Db 1381 acctgaagagcctgtgagggagagctggtcctgggggctcgaccggcgacactaca 1440  
QY 1441 tccacacctcaacttctgctccagtcacgtccctgctactgagatccacatggagc 1500  
Db 1441 tccacacctcaacttctgctccagtcacgtccctgctactgagatccacatggagc 1500  
QY 1501 gtgtgaaggtggggccagaggtgactctctgtgccaaggtgctgtgcccactcctgata 1560  
Db 1501 gtgtgaaggtggggccagaggtgactctctgtgccaaggtgctgtgcccactcctgata 1560  
QY 1561 cgggcaagtcctcactcaacagggaccactgagagatccggggccctgcatgagccattg 1620  
Db 1561 cgggcaagtcctcactcaacagggaccactgagagatccggggccctgcatgagccattg 1620  
QY 1621 ggggaatcccttgccttcttgggggggtgtggttaacctcaacgcccactgattacgtca 1680  
Db 1621 ggggaatcccttgccttcttgggggggtgtggttaacctcaacgcccactgattacgtca 1680  
QY 1681 ccgagctctcttcttcttgggggggtgtggttaacctcaacgcccactgattacgtca 1740  
Db 1681 ccgagctctcttcttcttgggggggtgtggttaacctcaacgcccactgattacgtca 1740  
QY 1741 tccagactactcgaaatggcgctccgctctgttgcgggtttccaggcccttgatgctc 1800  
Db 1741 tccagactactcgaaatggcgctccgctctgttgcgggtttccaggcccttgatgctc 1800  
QY 1801 ctccgctcagggcccttctgactcctcggtgacgtcttcttgggagcgtatgtggcg 1860  
Db 1801 ctccgctcagggcccttctgactcctcggtgacgtcttcttgggagcgtatgtggcg 1860  
QY 1861 tcttcgacccgggggacatgaagagcagcccggttggtggtcgtgcgctgcgactc 1920  
Db 1861 tcttcgacccgggggacatgaagagcagcccggttggtggtcgtgcgctgcgactc 1920



Db 841 tgaggaggatgagagagacagagctccccaaagtggggcccccattccccctggggaca 900  
QY 901 agccattctctacatctctcgaactacagggatgctcagtatcttggggaattggc 960  
Db 901 agccattctctacatctctcgaactacagggatgctcagtatcttggggaattggc 960  
QY 961 tgggaacgctccacaaactcactgttgcttgaactggctcctcaatctctgg 1020  
Db 961 tgggaacgctccacaaactcactgttgcttgaactggctcctcaatctctgg 1020  
QY 1021 tcccgccaggagatgcaactctcagtgctccctgctcgttaccacacagattgatc 1080  
Db 1021 tcccgccaggagatgcaactctcagtgctccctgctcgttaccacacagattgatc 1080  
QY 1081 ccaagcctctagctcctccaggccaatgggaccaagtggccattcaatgaactg 1140  
Db 1081 ccaagcctctagctcctccaggccaatgggaccaagtggccattcaatgaactg 1140  
QY 1141 ggcggtagatgaatcctgacgaggaacaagctactattggtgaatcaagggtgat 1200  
Db 1141 ggcggtagatgaatcctgacgaggaacaagctactattggtgaatcaagggtgat 1200  
QY 1201 cagtgatttccggaggctctctggagccagcctgcttcttccattttg 1260  
Db 1201 cagtgatttccggaggctctctggagccagcctgcttcttccattttg 1260  
QY 1261 atgggatatggcctcggtttccattctgtctgtggaaggagttcgcccgatgg 1320  
Db 1261 atgggatatggcctcggtttccattctgtctgtggaaggagttcgcccgatgg 1320  
QY 1321 atgtactggtgagcagggctattgataagcctgcttcttccattttacctcaacagg 1380  
Db 1321 atgtactggtgagcagggctattgataagcctgcttcttccattttacctcaacagg 1380  
QY 1381 accctgaagagcctgatgagagagctggtcctggtggggctcggaaccggcacaactaca 1440  
Db 1381 accctgaagagcctgatgagagagctggtcctggtggggctcggaaccggcacaactaca 1440  
QY 1441 tccacccctcacctctgacgtacaggtccctgctactgacagatccacatgagc 1500  
Db 1441 tccacccctcacctctgacgtacaggtccctgctactgacagatccacatgagc 1500  
QY 1501 gtgtgaagtggcccgaggtgactctctgtgcaaggctgtgctgcccactcgtgata 1560  
Db 1501 gtgtgaagtggcccgaggtgactctctgtgcaaggctgtgctgcccactcgtgata 1560  
QY 1561 cgggcacgtccctcacaacagaccactgaggagatccgggcccctgcatgagccattg 1620  
Db 1561 cgggcacgtccctcacaacagaccactgaggagatccgggcccctgcatgagccattg 1620  
QY 1621 ggggaatcccttctgctggtggagtagatcatcctctgctcggaatcccaagctcc 1680  
Db 1621 ggggaatcccttctgctggtggagtagatcatcctctgctcggaatcccaagctcc 1680  
QY 1681 ccgagctctcctcttcttgggggtctgtgttaacctcaagcccatgattacgtca 1740  
Db 1681 ccgagctctcctcttcttgggggtctgtgttaacctcaagcccatgattacgtca 1740  
QY 1741 tccagactactcgaatggcgctccgctctgcttctgctgttccaggcccctgagatcc 1800  
Db 1741 tccagactactcgaatggcgctccgctctgcttctgctgttccaggcccctgagatcc 1800  
QY 1801 ctccgctcgcagggccctcttgatctcgttgagctcttcttgggacagtattggcgg 1860  
Db 1801 ctccgctcgcagggccctcttgatctcgttgagctcttcttgggacagtattggcgg 1860  
QY 1861 tcttcgaccgctgggacatgaagagcagcccggtgggctggcgcgctcgactc 1920  
Db 1861 tcttcgaccgctgggacatgaagagcagcccggtgggctggcgcgctcgactc 1920  
QY 1921 gcggagcagacctcgatgggagagactgcgagggcgaggttccccgggtgacgccccaa 1980  
Db 1921 gcggagcagacctcgatgggagagactgcgagggcgaggttccccgggtgacgccccaa 1980

QY 1981 gtgaagcgcacgcgcagcgggtggtcgcggagggtcctctactccccagtaaaatccactat 2040  
Db 1981 gtgaagcgcacgcgcagcgggtggtcgcggagggtcctctactccccagtaaaatccactat 2040  
QY 2041 ttccattgaaaaaataaaaaa 2060  
Db 2041 ttccattgaaaaaataaaaaa 2060

## RESULT 3

AAV27038  
ID AAV27038 standard; DNA; 1910 BP.

XX AC AAV27038;

XX DT 26-OCT-1998 (first entry)

XX DE Human napsin B cDNA.

XX KW Napsin B; splicing; clone; screening; human liver cDNA library;  
aspartic protease; N-terminal; C-terminal; genomic clone; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 526..1788

XX FT /\*tag= a

XX FT /product= "napsin B protein"

XX PN WO9822597-A2.

XX PD 28-MAY-1998.

XX PF 20-NOV-1997; 97WO-US21684.

XX PR 09-MAY-1997; 97US-0046126.

XX PR 20-NOV-1996; 96US-0031196.

XX PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PI Keolsch G, Lin X, Tang JJN;

XX DR WPI; 1998-312482/27.

XX DR P-PSDB; AAW54878.

XX PT New isolated aspartic protease, napsin, from human liver -  
potentially useful for, e.g. diagnosis and treatment of disease

XX PS Claim 6; Figure 4; 24pp; English.

XX CC The present sequence represents the human napsin B cDNA. The N-terminus  
of this cDNA was obtained by splicing together isolated napsin cDNA  
clones, which had been found by screening a human liver cDNA library,  
whereas the C-terminus was obtained by using genomic clones. Napsin B is  
an aspartic protease which was isolated from human liver.

XX SQ Sequence 1910 BP; 437 A; 538 C; 522 G; 412 T; 1 other;

## Query Match

Best Local Similarity 65.9%; Score 1356.8; DB 19; Length 1910;  
Matches 1358; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 701 gtcccagcagatgtctccaccaccgctgtgaaacccctgctgctgctgctgctgct 760

Db 516 gtcccagcagatgtctccaccaccgctgtgaaacccctgctgctgctgctgctgct 575

QY 761 gaatgtggagcctccggggccaactgacatccctcttcacatcgagtcacactgg 820

Db 576 gaatgtggagcctccggggccaactgacatccctcttcacatcgagtcacactgg 635

QY 821 acgcaggaccctgaacctactgaggggatggagagaaccagcagatccccaagtgg 880

```
Db 636 acgaggaatcctgaacctactgagggatgagagaaacagacagagctccccaagtggg 695
Qy 881 ggcctccctcctgggacagcccatctctgacctctctcgaaactacagggatgtgca 940
Db 696 ggcctccctcctgggacagcccatctctgacctctctcgaaactacagggatgtgca 755
Qy 941 gtattttgggaaattggcctgggaacgctcccaaaacttcactgtgctttgacac 1000
Db 756 gtattttgggaaattggcctgggaacgctcccaaaacttcactgtgctttgacac 815
Qy 1001 tggctctccaatctctgggtcccgctcagagagatgccactttctcaagtgcctgtgctg 1060
Db 816 tggctctccaatctctgggtcccgctcagagagatgccactttctcaagtgcctgtgctg 875
Qy 1061 gtacaccaccgatttgatcccaagcctctagctctctccagcccaatggagacaaatt 1120
Db 876 gtacaccaccgatttgatcccaagcctctagctctctccagcccaatggagacaaatt 935
Qy 1121 tgccattcaatatggaactggcgggtagatggaaacctgagcagagacaagctgactat 1180
Db 936 tgccattcaatatggaactggcgggtagatggaaacctgagcagagacaagctgactat 995
Qy 1191 tgggtgaatacaagggtgcatcagtgatttttcggggaggctctctgaggccagcctgggt 1240
Db 996 tgggtgaatacaagggtgcatcagtgatttttcggggaggctctctgaggccagcctgggt 1055
Qy 1241 ctgcgcttttcccatttgatggagatattggcctcggttttccattctctgtctgtgga 1300
Db 1056 ctgcgcttttcccatttgatggagatattggcctcggttttccattctctgtctgtgga 1115
Qy 1301 aggagttcgcccccgcgatgtactgtgtgagcaggggctatttggaagcctgtctt 1360
Db 1116 aggagttcgcccccgcgatgtactgtgtgagcaggggctatttggaagcctgtctt 1175
Qy 1361 ctctttttacotcaacaggccctgaagcctgagatgagagcctgtggtctctggggggg 1420
Db 1176 ctctttttacotcaacaggccctgaagcctgagatgagagcctgtggtctctggggggg 1235
Qy 1421 ctggagcccgccacactacatccacccctcacctctgtccagtcacggtccctgccta 1480
Db 1236 ctggagcccgccacactacatccacccctcacctctgtccagtcacggtccctgccta 1295
Qy 1481 ctgagacatccacatggagcgtgtgaaagtgtggccagggtgactctctgtgccaaagg 1540
Db 1296 ctgagacatccacatggagcgtgtgaaagtgtggccagggtgactctctgtgccaaagg 1355
Qy 1541 ctgtgtgcacatcctgtgatacagggcacgtccctcatacagagcccaactgaggagatccg 1600
Db 1356 ctgtgtgcacatcctgtgatacagggcacgtccctcatacagagcccaactgaggagatccg 1415
Qy 1601 ggcctctgcacatgcacattgggggaatccctctgtcgtcgtggagtagacatcactctgtg 1660
Db 1416 ggcctctgcacatgcacattgggggaatccctctgtcgtcgtggagtagacatcactctgtg 1475
Qy 1661 ctgggaaatcccaagctcccccagctcctctccttcttgggggggtcgtggtttaacct 1720
Db 1476 ctgggaaatcccaagctcccccagctcctctccttctccttcttgggggggtcgtggtttaacct 1535
Qy 1721 cacggcccatgattacgttcatccagactactcgaataaggctcgcctctcgttgtccgg 1780
Db 1536 cacggcccatgattacgttcatccagactactcgaataaggctcgcctctcgttgtccgg 1595
Qy 1781 ttccaggcccttgatgtccctccgcctcagggccctcttgatccctcgtgacgtctt 1840
Db 1596 ttccaggcccttgatgtccctccgcctcagggccctcttgatccctcgtgacgtctt 1655
Qy 1841 ctgggggacgtatgtggcgtcttcgaccgcggggacatgaagagcagcgcgccgggtggg 1900
Db 1656 ctgggggacgtatgtggcgtcttcgaccgcggggacatgaagagcagcgcgccgggtggg 1715
Qy 1901 cctgggcgcgctcgcactcgcagagcgacctcgatggggagagactgcgcaggcgca 1960
|||||
|||||
```

```
Db 1716 cctggcgcgcctgcactcgcggagcggaacctcgatggggagagactgcgcaggcgca 1775
Qy 1961 gttccccgggtgacgcccgaagtgaagcagtcgcagcgggtggtcgcggaggctcctgct 2020
Db 1776 gttccccgggtgacgcccgaagtgaagcagtcgcagcgggtggtcgcggaggctcctgct 1835
Qy 2021 acccagtaaaatccactatttccatttgcatttgcataaaaaa 2060
Db 1836 acccagtaaaatccactatttccatttgcataaaaaa 1875

RESULT 4
AAZ50231
ID AAZ50231 standard; cDNA: 1263 BP.
XX
AC AAZ50231;
XX
DT 18-MAY-2000 (first entry)
XX
DE Human Aspartic Protease-1 (NHAP-1) encoding cDNA.
XX
KW Human; aspartic protease-1; NHAP-1; chromosome 19q13.3; antiallergic;
KW immunosuppressive; cytostatic; antiasthmatic; antinflammatory; cancer;
KW antiarteriosclerotic; antithyroid; antibacterial; neuroprotective;
KW antidiabetic; anti-HIV; osteopathic; antiarthritic; treatment; cretinism;
KW endocrinological; hypogonadism; Sheehan syndrome; diabetes insipidus;
KW hypothyroidism; adenocarcinoma; leukaemia; immunological; amyloidosis;
KW acquired immune deficiency syndrome; AIDS; Addison's disease; arthritis;
KW osteoporosis; atherosclerosis; infection; respiratory; allergy; asthma;
KW emphysema; gene therapy; diagnosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1263
FT /product= "Human Aspartic Protease-1 (NHAP-1)"
FT /note= "Homologous to mouse aspartic protease-like
FT protein (GI 1906810)"
FT sig_peptide 1..63
FT /tag= b
FT mat_peptide 64..1260
FT /tag= c
FT /product= "Mature Human Aspartic Protease-1 (NHAP-1)"
FT misc_binding 160..228
FT /tag= d
FT /bound_moiety= "Hybridisation probe"
XX
WO200004137-A1.
XX
PD 27-JAN-2000.
XX
PF 15-JUL-1999; 99WO-US15988.
XX
PR 16-JUL-1998; 98US-0116641.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Xu H, Bruno SA, Elsenboss LA, Fogliano M, Cohan VL, Bandman O;
XX
DR WPI; 2000-182413/16.
XX
DR P-PSDB; AAY44809.
XX
PT New human aspartic protease polypeptide useful for treating and
PT detecting endocrinological disorders e.g. hypogonadism, Sheehan
PT syndrome and diabetes insipidus.
XX
PS Example 6; Fig 1; 83pp; English.
XX
CC The present cDNA sequence encodes human aspartic protease-1 (NHAP-1).
CC It is obtained from human lung cDNA library. NHAP-1 gene is located on
CC chromosome 19q13.3. NHAP has immunosuppressive, antiallergic,
CC cytostatic, antiasthmatic, antinflammatory, antiarteriosclerotic,
```





Db	782	ctctactgctgagatccacatgagcgtgtgaagggtggcccgaggctga	ctctctctgtgcc	841
Qy	1537	agggctgtgctgccatctctggatacggcagtcctcatcacaggaccac	taaggaga	1596
Db	842	agggctgtgctgccatctctggatacggcagtcctcatcacaggaccac	taaggaga	901
Qy	1597	tccgggcctcgatgcagccattgggggaatccccttgcctggggagta	catcatcc	1656
Db	902	tccgggcctcgatgcagccattgggggaatccccttgcctggggagta	catcatcc	961
Qy	1657	tgtgctgcgaaatcccaagctccccgcagtcctctctcttcttctt	tggggggctggtta	1716
Db	962	tatgctctcqaatacccaaaagctcccgcagtcctctcttcttctt	tgaggagatcggtta	1021

[illegible]

Df	1062	- - - - -	tcttctgggacgtatgtgcgcttcgcaccgcgagcacatagaagcagcccccggg       -tgacg
Qy	1837	tcttctgggacgtatgtgcgcttcgcaccgcgagcacatagaagcagcccccggg	1896
Df	1067	tcttctgggacgtatgtgcgcttcgcaccgcgagcacatagaagcagcccccggg	1126
Qy	1897	tgggcctggcgcgctcgactcgaggagcagacctcgattggggagagaactgcgcagg	1956
Df	1127	tgggcctggcgcgctcgactcgaggagcagacctcgattggggagagaactgcgcagg	1186
Qy	1957	cgcagtccccgggtgaagcccaagtgaagcgcacatgcgcagcgggtggtgcgagagttcc	2016
Df	1187	cgcagtccccgggtgacgccccaagtgaagcgcacatgcgcagcgggtggtgcgagagttcc	1246
Qy	2017	tgtaccacagtaaaatccactatttcattgaaaaaaa	2060
Df	1247	tgtaccacagtaaaatccactatttcattgaaaaaaa	1290
RESULT 6			
ID	AHV28623		
AC	AHV28623 standard; cDNA; 1375 BP.		
XX	AHV28623:		
XX	(first entry)		
DE	Human aspartic protease encoding cDNA SEQ ID NO:2.		
KW	Human; aspartic protease; EST; endothelin; serum amyloid A protein; pro-opiomelanocortin prohormone; ss.		
OS	Homo sapiens.		
FH	Key Location/Qualifiers		
CDS	26..1339		
FT	/tag= a		
FT	/product= "aspartic protease"		
FT	/transl_except= (pos:1322..1324, aa:Arg)		
FT	/transl_except= (pos:1325..1327, aa:Ala)		
FT	/transl_except= (pos:1328..1330, aa:Ser)		
FT	/transl_except= (pos:1331..1333, aa:Ala)		
FT	/transl_except= (pos:1334..1336, aa:Thr)		
FT	/transl_except= (pos:1337..1339, aa:Gln)		
FT	/note= "no stop codon given; the CDS is marked according to the protein given in the specification but could decode for its full length"		
PX	WO9811236-A1.		
PD	19-MAR-1998.		
FF	09-SEP-1997;		
PF	97WO-GB02426.		

RESULT	6	
AAY28623		
ID	AAY28623 standard; cDNA; 137	
XX		
XX		
XX	AAY28623;	
XX		
DT	29-JUL-1998 (first entry)	
XX		
DE	Human aspartic protease enco	
XX		
KW	Human; aspartic protease; ES	
KW	pro-opiomelanocortin prohor	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qual
FT	CDS	26..1339
FT		/*tag= a
FT		/product= "a
FT		/transl_exce
FT		/transl_exce
FT		/transl_exce
FT		/transl_exce
FT		/transl_exce
FT		/note= "no s
FT		to t
FT		coul
XX		
PX	WO9811236-A1.	
XX		
PD	19-MAR-1998.	
XX		
PF	09-SEP-1997;	97WO-GB02426.





```

XX OS Homo sapiens.
XX FH Key
XX FT mat_peptide
XX FT 31..1263
XX FT /tag= a
XX FT /product= "Human Aspartic Protease-2 (NHAP-2)"
XX FT /note= "Homologous to mouse aspartic protease-like
XX FT protein (GI 1906810)"
XX FT 31..93
XX FT /tag= b
XX FT 94..1263
XX FT /tag= c
XX FT /product= "Mature Human Aspartic Protease-2 (NHAP-2)"
XX FT 190..258
XX FT /tag= d
XX FT /bound_moiety= "Hybridisation probe"
XX PN WO200004137-A1.
XX PD 27-JAN-2000.
XX XX
XX PF 15-JUL-1999; 99WO-US15988.
XX XX
XX PR 16-JUL-1998; 98US-0116641.
XX XX
XX PA (INCY-) INCYTE PHARM INC.
XX PI Xu H, Bruno SA, Elsenboss LA, Fogliano M, Cohan VL, Bandman O;
XX XX
XX DR WPI; 2000-182413/16.
XX DR P-PSDB; AAY44810.
XX PT New human aspartic protease polypeptide useful for treating and
XX PT detecting endocrinological disorders e.g. hypogonadism, Sheehan
XX PT syndrome and diabetes insipidus.
XX PS Claim 7; Fig 2; 83pp; English.
XX CC The present cDNA sequence encodes human aspartic protease-2 (NHAP-2). It
XX CC is obtained from human leucocyte cDNA library. NHAP-2 gene is located on
XX CC chromosome 19q13.3. NHAP has immunosuppressive, antiallergic,
XX CC cytostatic, antiasthmatic, antiinflammatory, antiarteriosclerotic,
XX CC antithyroid, antibacterial, neuroprotective, antidiabetic, anti-HIV,
XX CC osteopathic and antiarthritic activity. It is useful for treating and
XX CC preventing endocrinological disorders like hypogonadism, Sheehan
XX CC syndrome, diabetes insipidus, cretinism and hypothyroidism, cancers
XX CC like adenocarcinoma and leukaemia, immunological disorders like acquired
XX CC immune deficiency syndrome (AIDS), Addison's disease, amyloidosis,
XX CC arthritis, osteoporosis, atherosclerosis and microbial infections and
XX CC respiratory disorders like allergy, asthma and emphysema. NHAP
XX CC polynucleotides can be used in gene therapy and for diagnosis of
XX CC disorders associated with expression of NHAP.
XX SQ Sequence 1329 BP; 232 A; 408 C; 394 G; 295 T; 0 other;

Query Match 54.5%; Score 1122.6; DB 21; Length 1329;
Best Local Similarity 90.9%; Pred. NO. 0;
Matches 1194; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 697 acgcgtccacgcagatgtctccaccacgcgtgctgcaacccctgcgtgctgctgctcctc 756
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 757 tgcgtgaatggagcttcctcgggccacactgacatgcgcgtacccctctgctgctgctc 816
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 817 ctggacgcagaccctgaacctactaggggatggagaaacacagacagctcccaagt 876
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 137 ctgagcagcagaccctgaacctactagggggaagggaacacacagcagagctcccaagt 196

```

```

QY 877 tgggggccccatccccctgggggacaagcccatctctgtacctctctcgaactacaggatg 936
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 937 tgcagtattttggggaaattgggctgggaacgcctccacaaaacttcaactgttgcctttg 996
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 997 acctgctctccaatctctccgctccaggaagatgccacttctcaagtgcctt 1056
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1057 gctggttacaccacccgatttgatcccaaaacctctagctctctccaggaacaaatggacca 1116
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1117 agtttgccattcaatatgggaactgggcgggtagatgaaactcctgagcgagacaagtga 1176
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1177 ctattgtggaatcaagggtgcacatcagtgattttcggggaggtcctctggagcccagcc 1236
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1237 tggctcttgcctttgcccatttttgatgggatatattggccctcggttttccattctgtctg 1296
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1297 tgaagagatcgcccccgcgatgagtactggtggaagaggggctatttgataagcctg 1356
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1357 tcttctctttactcaacagggaccctgaagacgcctgagtggagagagctggtctgg 1416
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1417 ggggctcggaacccgcacactacatccaccctcactctctgctgcagtcacggctcctg 1476
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1477 cctactggcagatccacatggagcgtgtgaagtgggccagggctgactctctgacca 1536
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1537 agggctgtgctgcctccctggatcacgggacacgcctccctcatcacagaccactgaggaga 1596
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1597 tccggccctgcatgagcagcattgggggaatcccttgcctgctggctggggagatcatcctc 1656
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1657 tggctcggaatcccaaaagctcccgcagctcctctctcttggggggtctggttta 1716
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1717 acctacggcccatgattacgttoatccagactactcgaatggcgtccgcctctgctgtgt 1776
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1777 ccggtttccagaccctggatgctccctccctgcagggccctcttgatccctggtgacg 1836
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1837 tcttttggggacgtatgtggccgtcttcgaccgggggacatgaagacagcagcccggg 1896
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1897 tgggctgcgcgcgtcgcactcgcggagcgggaacctcggatggggagagactgcgcagg 1956
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1957 cgcagtccccggggtagacgccccaaagtgaagcgcatgctgcagcgggtggtcgcg 2009

```



QY 1770 tctgttcggtttccaggccctgagtgctccctcgcctgcagggccctcttgatctc 1829  
 Db 1123 tctgttcggtttccaggccctgagatcgctcggtggaaggacctgtctgatactc 1182  
 QY 1830 ggtgacgtcttttgggagcgtatgtgctcttcgcacccgagacatgaagagcgc 1889  
 Db 1183 ggcgaagtttttgggaa--tatgtgaccttttgacgggggggacatgaagagcgga 1240  
 QY 1890 gcccggtgggctgcgcgcg 1909  
 Db 1241 acccgagttgacttgcgggg 1260

## RESULT 10

AAZ94089

ID AAZ94089 standard; cDNA; 1360 BP.

XX AC

XX AAZ94089;

XX 19-JUN-2000 (first entry)

XX Haematopoietic stem cell specific nucleic acid.

XX B4-14; haematopoietic stem cell; immune system disorder;

XX leukaemia; antileukaemic; immunomodulator; therapy; mouse; ss.

XX OS Mus musculus.

XX PN W0200011168-A2.

XX XX 02-MAR-2000.

XX XX 20-AUG-1999; 99WO-US19052.

XX XX 21-AUG-1998; 98US-0138132.

XX XX (UYPR-) UNIV PRINCETON.

XX XX Lemischka I, Moore K;

XX XX WPI; 2000-237650/20.

XX Hematopoietic stem cell signaling proteins modulating replication and  
 XX differentiation for treating immune system disorders and leukaemia -  
 XX Claim 10; Page 203; 256pp; English.

CC The present sequence is that of a nucleic acid isolated from  
 CC mouse primitive stem cells by methods of the invention. It is an  
 CC example of claimed isolated nucleic acids (see AAZ94077-294131) that  
 CC which are specifically expressed in haematopoietic stem cells (HSCs) and  
 CC primitive HSCs (PHSCs) such as umbilical cord cells, bone marrow  
 CC cells and foetal liver cells. The encoded proteins (see AAY79176-93)  
 CC are growth factors, transcription factors, splicing factors, capping  
 CC factors, transport proteins, translation factors or replication  
 CC factors that modulate HSC activity, especially differentiation or  
 CC replication. The invention provides a claimed method for  
 CC identifying PHSC-specific nucleic acids involving: creating a PHSC  
 CC cDNA library and a non-PHSC immune cell library; and subtracting  
 CC the 2 libraries. Also claimed are methods for generating a stem  
 CC cell/progenitor cell from PHSCs; for identifying the presence of a  
 CC PHSC in a sample; for identifying the presence in a sample of a  
 CC compound that modulates HSC activity; for using such a compound to  
 CC treat an immune system condition, especially leukaemia, for  
 CC introducing exogenous nucleic acid into a HSC; and for ex vivo  
 CC expansion of HSCs. Also claimed is a PHSC specifically expressing 1  
 CC of the claimed nucleic acids, such as the present sequence.

SQ Sequence 1360 BP; 293 A; 385 C; 355 G; 327 T; 0 other;

Query Match

38.7%; Score 797.2; DB 21; Length 1360;

Best Local Similarity 75.6%; Pred. No. 5.1e-211;  
 Matches 1034; Conservative 0; Mismatches 318; Indels 16; Gaps 3;

QY 694 agaagcggtcccccagcagatgtctccacacacccgctgtgcacacccctgctgctgctgcgc 753

Db 1 agctctagtcgccagagatgtcgcacactactgctg-----ctgctgctgctgcgc 51

QY 754 ctctgctgaatgtgagccttccggggccacacactgacgcgcatccctcttcatcgagtc 813

Db 52 tgcctggggaaatttgagcctgagggagggccaaactgacctgtccctcttcaacgaatcc 111

QY 814 aacctgacgcagcagcaccctgaacctactgaggggatgagagaaacacagcagagctcccca 873

Db 112 accttgacacagaatcttaaacccactgaatggatgggaacagctggcgagagcttcta 171

QY 874 agttgggggccccatccccctggggacaaagccatctctctgacctctctcgaactacaggg 933

Db 172 -----ggacctccacctctgtgtgcaacccctcttctgctctctccaagtcatga 225

QY 934 atgtgcagtatatttgggaaattggcgtgggaacgcctccacaaaacttcaactgttgcct 993

Db 226 acaccagtatatttggaaactattgtttgggaacgcctctcagaatttcacogttgct 285

QY 994 ttgacactggctctcccaatctctgtgctccctccagagatgccactcttctcagttgc 1053

Db 286 ttgacacgggttctcccaacttgtgggtccgtccacagagatgcatttctcagttgg 345

QY 1054 cctgctgtttacacacccagatttgatcccaagcctctagctctctccagcccaatggga 1113

Db 346 catgctggtttccacctcgctttaaaccgaagcctccagctctcaggcccaatggga 405

QY 1114 ccaagttgcatcattcaatatggaactggcggtgagatggaatccctgagcagagacaagc 1173

Db 406 ccaagtttgcattcagtatgggaacggcgctgagcggaaacctgagcaggagaatc 465

QY 1174 tgactattgtggaatacaagggtgcacatcagtgattttcggggaggctctctggagccca 1233

Db 466 tgactatcgggggatccacagatgctttgtgacatttgagaggctctgtggagccca 525

QY 1234 gctggtcttcgcttttgcatttgatgggatatttgacctggtctcttccactctgt 1293

Db 526 gctgactcttgccttagccactttgagggatcctggctcggctcggctctcccaacttg 585

QY 1294 ctgtgaaaggagttcggcccccgcgatggtactggtgagcaggggctattggataaagc 1353

Db 586 ctgtggcgagggttcagctccgctggtgagtgatggtgagcaaggctgctggagaaac 645

QY 1354 ctgtctctctcttttacctcaacaggggacccctgaagagcctgatggagagctggtcc 1413

Db 646 cgtgtctctctcttttacctcaacagggttctgaagggtctgatggggagagctggtcc 705

QY 1414 tggggggctcggacccggcacactacatccaccctcactctggtgcagtcagcggctcc 1473

Db 706 tagggggctcagaccccgctcactacgtacctccctcactctcaccagtcacatcc 765

QY 1474 ctgctactggcagatccacatggagcgtgtgaaagtggtggccagggtgactctctgtg 1533

Db 766 ctgctactggcaggtccacatggagagtgtaagggtggcagagggtgagcctctgtg 825

QY 1534 ccaagggtctgtgctccatccctggatacgggcacgtccctcactcaacaggaccactgag 1593

Db 826 ccaagggtctgagtgccatccctgacacaggcacatccctcactcaacaggaccactgag 885

QY 1594 agatccgggcccctgcatgcagcattggggggaatcccccttgcgtggtgggagtagatca 1653

Db 886 agatccgggccccttgataaagcattgggggatacccccttgcgtggtgggagtagatca 945

QY 1654 tctgtgctcggaatacccaagctccccgagtcctctcttcttcttgggggggctggt 1713

Db 946 ttcagttcccaagagcgccacgcttccccctgtctctcttccaccttgggtgagctggt 1005

QY 1714 ttaacctcacgcccctgatgattacgtcatccagactactcgaatggcgtccgctctgct 1773









## RESULT 13

```
AAFI8208
ID AAF18208 standard; DNA; 1041 BP.
XX
AC AAF18208;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 227.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
WO200005180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-0505918.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
WPI: 2000-587514/55.
DR P-PSDB; AAB58332.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer
XX
PS Claim 1; Page 689; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1041 BP; 227 A; 313 C; 282 G; 219 T; 0 other;

Query Match          29.2%; Score 602.4; DB 21; Length 1041;
Best Local Similarity 84.6%; Pred. No. 4.7e-157;
Matches 790; Conservative 0; Mismatches 16; Indels 128; Gaps 5;

QY 693 cagaacgcgtccacgcgagtgtctccaccacgcgtgtgtgcaaccctgtgctgctgtg 752
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 cacaccacgtccacgcgagtgtctccaccacgcgtgtgt-aaaccctgtgctgctgtg 294
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 753 cctctgtgaatgtggagccttcggggccacacgtgacccgcatccctcttcatcagtc 812
```

```
Db 295 cctctgtgaatgtgga-ccttcggggccacactgacgcacatcctcttcacgagtc 353
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 813 caacctggacgcagcagccctgaacctactgaggggagtgagagaaaccacgagctcccc 872
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 354 caacctggacgcagcagcctgaacctactgaggggagtgagagaaaccacgagctcccc 413
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 873 aagttggggcccccattccctctgggggacaagcccatctctgtactctctcgaactacaag 932
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 414 aagttggggcccccattccctctgagggacaagcccatctctgtactctctcgaactacaag 473
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 933 gatgtcagttatttggggaatttgggtgggaacccctccacaaacttcaactgttgc 992
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 474 g----- 474
QY 993 ttgacactgggtccctccaatctctgttcccggtccaggagatgccacttcttcagtgtg 1052
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 475 ----- 474
QY 1053 cctgtgtgttacaccacccagatttgatcccaagcctctagctcctccaggccaatggg 1112
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 475 ----gatggttacaccacccagatttgatcccaagcctcta--ctccttccagccaatggg 528
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1113 accaagtttgcattcaatataatggaaactggcggttagatggaaatcctgacgcaggaacaag 1172
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 529 accaa-tttgccattcaatataatggaaactggcggttagatggaaatcctgacgcaggaacaag 587
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1173 ctgactatttggtagaatacaagggtgcacagtgattttcggggagagctctcttggagccc 1232
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 588 ctgactatttggtagaatacaagggtgcacagtgattttcggggagagctctcttggagccc 647
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1233 agcctgtctctgtcttggccattttgatgggataattgggctcgtgtttccattctg 1292
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 648 agcctgtctctgtcttggccattttgatgggataattgggctcgtgtttccattctg 707
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1293 tctgtggaaggagttcggcccccagtgatgtactgtgtgagcaggggctattggataag 1352
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 708 tctgtggaaggagttcggcccccagtgatgtactgtgtgagcaggggctattggataag 767
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1353 cctgtctctctcttctacccaacaggggaccctgaagagcctgagagcctgagagagctggct 1412
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 768 cctgtctctctcttctacccaacaggggaccctgaagagcctgagagcctgagagagctggct 827
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1413 ctggggggctcgaccgcggcacactacatcccccactcctcctcctcctcctcctcctcctcct 1472
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 828 ctggggggctcgaccgcggcacactacatcccccactcctcctcctcctcctcctcctcctcct 887
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1473 cctgcctactgcagatccacatcagcagcgtgtgaaagtggccagggctgactcctcctg 1532
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 888 cctgcctactgcagatccacatcagcagcgtgtgaaagtggccagggctgactcctcctg 947
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1533 gccaaaggctgtgctgctgccatcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1592
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 948 gccaaaggctgtgctgccatcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1607
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1593 gagatccggggcctcgtcagcagccattgggggaa 1626
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1008 gagatccggggcctcgtcagcagccattgggggaa 1041
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
AAV28634
ID AAV28634 standard; cDNA; 481 BP.
XX
AC AAV28634;
XX
DT 29-JUL-1998 (first entry)
XX
DE Human aspartic protease partial cDNA EST 1320439 SEQ ID NO:12.
XX
KW Human; aspartic protease; EST; endothelin; serum amyloid A protein;
KW pro-opiomelanocortin prohormone; ss.
XX
```

OS Homo sapiens.  
XX WO9811236-A1.  
XX PD 19-MAR-1998.  
XX PF 09-SEP-1997; 97WO-GB02426.  
XX PR 11-SEP-1996; 96GB-0018966.  
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX PA (UYWA-) UNIV WALES.  
XX PI Hill J, Kay J, Powell D;  
XX WPI; 1998-207396/18.  
XX Human aspartic protease and related DNA - which may be involved in  
XX processing of endothelin and pro-opiomelanocortin pro-hormone(s)  
XX Disclosure; Page 22; 32pp; English.  
XX  
CC The present sequence represents a partial cDNA sequence from a human  
CC aspartic protease, isolated from a cDNA library of human origin. The  
CC present invention describes novel human aspartic proteases and also ESTs  
CC from human cDNA libraries having partial DNA sequences which encodes an  
CC aspartic protease. Compounds which inhibit aspartic protease, especially  
CC antibodies can be used in therapy where needed. The protease can also be  
CC used in therapy where needed. Important functions of aspartic proteases  
CC are the processing of endothelin and pro-opiomelanocortin prohormones.  
CC They may also be involved in the processing of serum amyloid A protein.  
XX  
SQ Sequence 481 BP; 90 A; 132 C; 135 G; 124 T; 0 other;

Query Match 19.3%; Score 397.2; DB 19; Length 481;  
Best Local Similarity 92.6%; Pred. No. 2.6e-100;  
Matches 428; Conservative 0; Mismatches 33; Indels 1; Gaps 1;  
QY 957 gggctgggaacgcctcccaaaactcactgttgccttgacactggctccctcaatctc 1016  
DB 20 gggctgggaacgcctcccaaaactcactgttgccttgacactggctccctcaatctc 79  
QY 1017 tgggtcccgctcaggagatgccactttctcagtgccctgctggttacaccacgattt 1076  
DB 80 tgggtcccgctcaggagatgccactttctcagtgccctgctggttacaccacgattt 139  
QY 1077 gatcccaagcctctagctctcctcaggcccaatgggaccacagtttgccattcaatgga 1136  
DB 140 aatcccaatgccctcagctcctcctcaagccagtgggaccacagtttgccattcaatgga 199  
QY 1137 actggcggttagatggaatcctgagcagacagctgactatggtggaatcaagggt 1196  
DB 200 actggcggttagatggaatcctgagcagacagctgactatggtggaatcaagggt 259  
QY 1197 gcatcagtgattttcgggagagctctcgtggagccagcctgctgtcttgcctat 1256  
DB 260 gcatcagtgattttcgggagagctctggtgaatccagcctgctgtcttgcctat 319  
QY 1257 ttgtatgggatatgggctcgtgtttccattctctgtgtggaaggagttcgcccccg 1316  
DB 320 ccgatgggatatgggctcgtgtttccattctgctgtggaaggagttcgcccccg 379  
QY 1317 atggatgactgtgtagcagggctattgataagcctgtctcttcttacctcaac 1376  
DB 380 cggatgactgtgtagcagggctattgataagcctgtctcttcttacctcaac 439  
QY 1377 agggaccctg-aagagcctgtagggagagagctggtctctggg 1417  
DB 440 agggaccctgaaagtggctgattgaagagaaactgtctctggg 481

RESULT 15

AAC00579  
XX ID AAC00579 standard; cDNA; 407 BP.  
XX AC AAC00579;  
XX DT 06-OCT-2000 (first entry)  
XX DE Human secreted protein 5' EST, SEQ ID NO: 577.  
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping; ss.  
XX OS Homo sapiens.  
XX PN EP1033401-A2.  
XX PD 06-SEP-2000.  
XX PF 21-FEB-2000; 2000EP-0200610.  
XX PR 26-FEB-1999; 99US-0122487.  
XX PA (GSEST ) GENSET.  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX P-PSDB; AAG00573.  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 577; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
SQ Sequence 407 BP; 79 A; 137 C; 97 G; 92 T; 2 other;

Query Match 18.8%; Score 387.6; DB 21; Length 407;  
Best Local Similarity 99.0%; Pred. No. 1.1e-97;  
Matches 398; Conservative 2; Mismatches 1; Indels 1; Gaps 1;  
QY 701 gtccccagcgatgtctccacaccgctgtgcaacccctgctgctgctgctgctgtgt 760  
DB 7 gtccccagcgatgtctccacaccgctgtgcaacccctgctgctgctgctgctgtgt 66  
QY 761 gaatgtgagccttcggggccacactgacatccctcttcctcatcgatccaaactgg 820  
DB 67 gaatgtgagccttcggggccacactgacatccctcttcctcatcgatccaaactgg 126  
QY 821 acgcaggaccctgaacctactgaggggatggagagaacacagcagagctcccaagtggg 880  
DB 127 acgcaggatcctgaacctactgaggggatggagagaacacagcagagctcccaagtggg 186  
QY 881 ggccccatccccctggggacaagccccattctctgactctctcgaactacaggatgca 940  
DB 187 ggccccatccccctggggacaagccccattctctgactctctcgaactacaggatgca 246

